

QY 753 TPMP-----PPTSEVSTPTTK-----EPTHHSPEDETPBELSAEPTPKALENSPK 801
 DB 1542 SREPTVPAPVPKPNPSHKSTKIELKEAPILPSPPTSPKIPSSKKAAPT--SAPK 1599
 QY 802 E-PEVPTTKPAATKPEKTTAKDTERDRLTPEPTTAAPKMTKEATTEKTESKI 860
 DB 1600 EFPASPSIK-PVTT--SLAQTAPPSLOKAPSTTIKEMALAPAV-----LPVSSKSPAPA 1652
 QY 861 TATTOVSTSTTQDTPPKITTLTKTTLAPKVTTKTKTTTTELMNPEETAKKDRATN 920
 DB 1553 RASASLSBPATAAPOTAKREATTIPSCKAATEPIETSTAPSLGAPKETSE-----TSV 1708
 QY 921 SKATTPPKPOTKAPKPKPTSTKPKPTPRVKRPTKPTPKMTSTMDLNPSTRIAEAML 980
 DB 1709 SKVLMSSP-----PKASSSKRASTLP-----ATTLSLKRASVLS----- 1744
 QY 981 QTTTRPOTRPSKIVENPKSESDAGAGETP 1012
 DB 1745 -----PTATSSGNDSHSPVS-DACSTGTTP 1770
 RESULT 14
 Q9VEL9 PRELIMINARY; PRT; 2112 AA.
 ID Q9VEL9
 AC Q9VEL9
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE CG4090 PROTEIN.
 GN CG4090.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amandites P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Abpayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houch J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibergam C.,
 RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mlshina N.V., Modyar C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
 RA Rehert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden Klamis I., Simpson M., Skupski M.P., Smith T.,
 RA Spiers R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaverl J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003716; AAF55402.1;
 DR FLYBASE: FBgn0038492; CG4090.
 DR InterPro: IPR002557; Chitin binding.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF01607; Chitin_bind_2; 10.
 DR SMART: SM00494; ChEBD2; 11.
 DR PROSITE: PS00022; EGF_1; 1.
 SQ SEQUENCE 2112 AA; 219547 MW; B91018E5551A5D36 CRC64;
 Query Match 10.6%; Score 752; DB 5; Length 2112;
 Best Local Similarity 26.6%; Pred. No. 3.3e-42;
 Matches 289; Conservative 105; Mismatches 356; Indels 338; Gaps 47;
 QY 78 FERGRCDCD-----AOCKKYDKC-----CPDESPCAEYKD-NKKNRT 115
 DB 969 FKPAKCESEETFLADENCKFRVCYDNGKGFVKVSTCPPTLMDPEANSCNHPDQI 1028
 QY 116 KKKPVPVVDENAGSLDNGDFKVTTPDTSTTOHKNVSTSPKITAKPTNPRESLPPNS 175
 DB 1029 QTKPLKCKKVVSGGSSSN-----STSNSSSSSSNNGSSSSSSSSSSSS 1076
 QY 176 DTSEETSLYVKKETV---ETKETTNNKOTS---TDGEKETSARETOSI---EK 222
 DB 1077 GSSSSNTSSSSNSGASGSSGSSSSGSSSSGSSSSGSSSSGSSSSSSSSSSSSNNNOG 1136
 QY 223 TSAKDLPTSKVLAKPPPAKETTTPG----- 249
 DB 1137 SSSSSSSSSSSSTSKRPSETCKVNGOPIGDRSDCAFYCVNDRCGFMMFPSCGPGT 1196
 QY 250 -----ALTPKEPTPTPKPEASTPKPEPTPTTKKAPPTPKPEAP 290
 DB 1197 VMDAQMOCNHMAVAKCEGCIAPPTTPTTSR-PTTASTSRSDQSTSRPTGP---P 1251
 QY 291 TTKKSATPTTKKPAPTTKKPAPTTKKPAPT- 323
 DB 1252 TTAAPVTARPTSSPTTASSQSTSPVTOAPNTDGCRCSEGFMAPDNCKSKFYCVANNK 1311
 QY 324 ---TKEP---APT-----TTKSAPPTKPAPTTKKPAPTTKKPAPT 360
 DB 1312 GGTSTIPFOGAGVMDQDLOTGNHNNNSSTGTESTPKRPP--EPATNGTATSTSSST 1369
 QY 361 TTPKEPTPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKK 420
 DB 1370 TTP--PPTTDLPTTSGTGLP-PTTTTELPT- -TTTDLPTTTRLPPTTTTSL 1419
 QY 421 PSPPTTKKPAPTTKKSAPTTKKPAPTTKKSAPTTK--EPSPPTTKKPA----- 469
 DB 1420 PPTTTGLPPTTTGAOPTTTLSSEETSTVTTSSESTTPPSTTMKLPAGTECTGE 1479
 QY 470 -----TTPK-----TPAP---TTPKPPAP 485
 DB 1480 GYMADEDCKKYRCINAGASYKNYFCKGGMNEEVOYCYENIPKCSLPAPPT 1539
 QY 486 TTP-----KEPAPTTPKE-PAPTTKKPAPTKPAPTTKKPAPTTKKPAPTTKKPA 540
 DB 1540 TTPSESKDPSSTPOSTDEPTVTXPI-TKPE-EPSTKEPKOPTYXPEKPTTEE-- 1595
 QY 541 PTPKEKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKK 599
 DB 1596 PEKPKPT-TTEYPKPTTEPTPEKPKPT-TTEYPKPTTEPTPEKPKPTTEPTPEK 1653
 QY 600 PKKPAPTTKKPA-----TTPK--GTAPPTKPAPTTKKPAKPAKPAELAPPTTKKPTST 652
 DB 1654 PKKPTTEPTTSSIRPYNPTTTSVPKYNPTTTPIPETTTSTPGK---PTTGTGPTTT 1710
 QY 653 SSKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKK 706
 DB 1711 T-LPSTTTDAIOEPTTSKRKEPTTTTESPSSPEGSAVTTLQDEPOPNYNCSESGEFPDP 1769


```

QY 707 KEIA-----PTTKGPTSTSDKPAETPKETAATPKBAPATPKKPAETPEPT 760
Db 1770 EDCGRYRCVDAKNGKAYQYAFKCGKVTWDTSTECNVADQSGN-----CS 1818
QY 761 TSEVSTPTTKKEPTTKKSDESTEPLESAETPKALENSKPEGVPTTKPAATPEMT 820
Db 1819 SGQTTTGTTEPCTESTSSGNP-----ETTSKAPEN-----TTWA---PETTT 1862
QY 821 TAKOKTERDRLRTPETTTAARPKMTKETATTEKTESKITATTTQVTSSTTODTTPFKI 880
Db 1863 T-----SSEPTTTT---VASETTTTTSGT-----TTTAPEPTTKPKP--- 1898
QY 881 TTKTTTAAKVTYTKKTTTTEIMNKPETAKPKDRATNSKATTPKPKOKETAKPKPTS 940
Db 1899 ---ETTIIAGEITSTKSPTTTE-----SPAPSTWTSAP----- 1929
QY 941 TKPKTMP 948
Db 1930 --CPETGP 1935

RESULT 15
Q9XDH2 PRELIMINARY; PRT; 763 AA.
ID Q9XDH2
AC Q9XDH2;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE POLYLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Espitia C., Lacleite J.L., Mondragon-Palomino M., Martens A.,
RA Zhang Y., Moreno C., Singh M.;
RT "Cloning and characterization of a new member of the PGHS family that
RT is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF071081; AAD41594.1; -
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; P_rich_extensn.
DR PRINTS; PRO1222; ACTROPHIN.
DR PRINTS; PRO1217; PRICHEXTENSIN.
DR PRINTS; PRO1218; PSTLEXTENSIN.
SQ SEQUENCE 763 AA; 75034 MW; 39168EC45A5916F8 CRC64;

```

Query Match 10.1%; Score 715; DB 2; Length 763;
 Best Local Similarity 28.2%; Pred. No. 3.5e-40;
 Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

```

QY 238 PTPKAETTTKGPATTPKEPTTPKEPASTPKKEPTTTIKSAPTTPKEBAPTTPKTSAP 297
Db 3 PVP-----APRALAPLPAPAPAPAEKSKPPPPAP-----PAPFCMMLYSAP 46
QY 298 TTPKEBAPTTPKEBAPTTPKEBAPTTPKESAPTTPKEBAPTTPKPAETPK 357
Db 47 PCP---PAP-----PAPPKPKSKAPEPPVPAPAPARELAPLP---PAP-----PEAPRE 90
QY 358 PAPTPPKPEPTT---PKBAPTTPKEBAPTTPK-EPAPTAPK-KPAPTTPKEBAPTTPK 411
Db 91 SRPALPCCPPPPVVIIPDEPAAPVPAPAPNSPPPPPPAPKAVPAPVP---PVPNSP 148
QY 412 EP--APTTPKEBAPTTPKESAPTTPKEBAPTTPKESAPTTPKESAPTTPKEBAP 469
Db 149 FPPPPPALNPPAP-----PAPPLANSPLPPAPPTPAGT--PRAAPMPVPAPAKSKPA 201
QY 470 TTPKEBAPTTPKPAETTPKEBAPTTPKEBAP---TTTKKPAETAPKEBAPTTPKETAP 525

```

```

Db 202 SPPAPAP-----PAPATPMEFPPLPVVPDPDISKETTPAPAPAPIPAPVIPPVPLP 256
QY 526 TTPKLLPTTPPEKLAP-----TTPKEBAPTTPBEELAPTPPEPTTPPEBAPTTP-- 575
Db 257 PVPKIPPPAP--APVVAANAVALVAPCPPLPLPLPNHPPAPAPVPGVPLAPLPNSHP 313
QY 576 ---PKAANPKKEBAP----- 589
Db 314 PAPSAVPGVPLAPLPISGRVSVWKGSPFTLSTFCRCVSGEVLGALNPSRSP 373
QY 590 --TTPKEBAPTTPKEBAPTTPKETA---PTTPKGTA---PTTKEBAPTTP-KKPAKE 639
Db 374 TTTTAPLAPAPLPPLPLPLPINTAVPPIPLPLPVLTALAPLPAPLPISGVVAP-- 431
QY 640 LAPTTKEPSTSDKPAETPKGNAPTTPKEBAP-----PTTPKEBAPTTPKGAPTTL 693
Db 432 --PIPPGKWTTPPLAPAPPEPK-TVPVLPFGPCPSEKPNPAPAPPEPPPKSSPALP 488
QY 694 KEBAPTTP---KKPAKELAPTTTGGP-TSTSDKPAETTPKET--APTTPKEBAPTTP 746
Db 489 APAPASMSAVRVPSPPIPPAPAPASMAPALPAPAPSPATRLCPPLPSPAPAPNSP 548
QY 747 KKPAPTTPETPPTTSEVSTPTTKKEPTTIKSP---DESTPELSAETPKA--LENSP 800
Db 549 --PAPAPAPTTPKLLS--ANPPCPPVPPAPNRPAPAPAPAPAPPELPAPDPPTPVANS 604
QY 801 KEGVPTTKTPAAKPEMTTIAKDKTTERDLRTTETTAAPKMKTKETATTEKTESKI 860
Db 605 --PAPAPAPAPALPEVNPPA-----PPTPAAPK-----SRPAL 637
QY 861 TATTTQVSTTODTTPPKITTLKTTTLAPKVTTKKTTTTEIMNKPETAKPKDRATN 920
Db 638 PAPAPAPAPVRAATP-----PPAPAPAPN 665
QY 921 SKATTPKQKP---TKAPKPTSTKKPKTTPRVRKKPTTPPKKMTSM---PELNP 971
Db 666 SMALEPPAPDPPIPLATPPAPAPPLPMSPPAPPLPAAADPPAPPLTLINQPSPLAP 725
QY 972 TSRIAEMLQTTTRPNOTPNSKLVEVNPKSSEDAGGEGT 1011
Db 726 VPGAPLAPLPINGRVEARKNSLI-----GSSSGDT 756

```

Search completed: April 26, 2002, 16:30:35
 Job time: 650 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:29:54 ; Search time 210.7 Seconds
(without alignments)
944.835 Million cell updates/sec

Title: AA7
Perfect score: 7333
Sequence: 1 MAWKTLPIYLLLSVEFIQ.....ARAITTSGQTLSKWNCP 1361

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7291.5	99.6	1404	4 Q92954	Q92954 homo sapien
2	7276.5	99.4	1404	4 Q9BX49	Q9BX49 homo sapien
3	3826.5	52.3	1054	11 Q9JMG9	Q9JMG9 mus musculi
4	1713.5	23.4	401	6 Q77765	Q77765 bos taurus
5	982	13.4	1079	5 Q9NAS7	Q9NAS7 caenorhabdi
6	955	13.0	1049	5 Q917S1	Q917S1 drosophila
7	868.5	11.9	1795	5 Q76894	Q76894 drosophila
8	867	11.8	1489	10 Q96449	Q96449 phytophthor
9	839.5	11.5	1315	10 Q9SPM0	Q9SPM0 zea mays (m
10	835	11.4	1274	5 Q20007	Q20007 caenorhabdi
11	830	11.3	1480	10 Q9LIE8	Q9LIE8 arabidopsis
12	828.5	11.3	2187	11 P70670	P70670 mus musculi
13	800	10.9	1188	10 Q41805	Q41805 zea mays (m
14	757.5	10.3	2112	5 Q9VEI9	Q9VEI9 drosophila
15	715	9.8	763	2 Q9XDH2	Q9XDH2 mycobacteri
16	707.5	9.7	555	10 Q9FPQ6	Q9FPQ6 chlamydomon
17	706	9.6	6677	5 Q9NA35	Q9NA35 caenorhabdi
18	704.5	9.6	1151	13 Q57580	Q57580 gallus galli
19	688	9.4	4880	11 Q9JLTI	Q9JLTI ratius norv

20	688	9.4	5085	11 Q9UKS6	Q9UKS6 ratius norv
21	687.5	9.4	2284	5 Q9VPG1	Q9VPG1 drosophila
22	674	9.2	1229	5 Q94185	Q94185 caenorhabdi
23	670	9.1	7962	4 Q10465	Q10465 homo sapien
24	668	9.1	3507	5 Q23587	Q23587 caenorhabdi
25	663	9.1	3570	4 Q99552	Q99552 homo sapien
26	659.5	9.0	4833	11 Q9QYX6	Q9QYX6 mus musculi
27	659.5	9.0	5038	11 Q9QYX7	Q9QYX7 mus musculi
28	657.5	9.0	2089	4 Q14676	Q14676 homo sapien
29	654	8.9	2768	5 Q9VC00	Q9VC00 drosophila
30	653.5	8.9	1514	5 Q9GUM7	Q9GUM7 leishmania
31	646	8.8	990	13 Q91803	Q91803 xenopus lae
32	639	8.7	1612	5 Q9VYQ2	Q9VYQ2 drosophila
33	633	8.6	489	10 Q41707	Q41707 vigna ungu
34	633	8.6	6632	5 Q17362	Q17362 caenorhabdi
35	632	8.6	761	10 Q9Z010	Q9Z010 arabidopsis
36	632	8.6	971	5 Q9XYS4	Q9XYS4 caenorhabdi
37	626	8.5	6642	5 Q01761	Q01761 caenorhabdi
38	622.5	8.5	839	2 Q9RX57	Q9RX57 delinococcus
39	620.5	8.5	2344	5 Q9N3Y8	Q9N3Y8 caenorhabdi
40	607.5	8.3	801	5 Q23635	Q23635 caenorhabdi
41	607.5	8.3	924	12 Q99307	Q99307 epstein-bar
42	605	8.3	379	5 Q27929	Q27929 drosophila
43	600.5	8.2	1893	5 Q9NKC9	Q9NKC9 drosophila
44	598.5	8.2	409	10 Q9SBM1	Q9SBM1 volvox cart
45	577	7.9	956	10 Q9LJ64	Q9LJ64 arabidopsis

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	1404	AA.
ID	Q92954				
AC	Q92954;				
DT	01-FEB-1997 (TREMBLrel. 02, Created)				
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	MEGAKARYOCYTE STIMULATING FACTOR.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxId=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,				
RA	Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,				
RA	Berez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,				
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;				
RL	Blood 78:279-279(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Merberg D.M., Fitz L.J., Temple P., Giannotti J., Wurtha P.,				
RA	FitzGerald M., Scaltireto J., Kelleher K., Preissner K., Kriz R.,				
RA	Jacobs K., Turner K.;				
RL	(in) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,				
RL	Mosher D.F. (eds.);				
RL	Biology of Vitronectins and their receptors., pp.45-52,				
RL	Elsevier Science Publishers B.V. (1993).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,				
RA	Kelleher K., Giannotti J., Calvetti J., FitzGerald M., Kriz M.J.,				
RA	Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,				
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;				
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.				
RL	EMBL; U70136; AAB0989.1; -				
DR	InterPro: IPR000585; Hemopexin.				
DR	InterPro: IPR001212; Somatomedin_B.				
DR	InterPro: IPR002400; GF_cysknott.				
DR	Pfam: PF00045; hemopexin; 2.				
DR	Pfam: PF01033; Somatomedin_B; 2.				
DR	PRINTS: PR00438; GFCYSKNOT.				

DR PRINTS: PR00022; SOMATOMEDINB.
DR PROSITE: PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE: PS00524; SOMATOMEDIN_B; 2.
DR SMART: SM00120; HX; 2.
DR SMART: SM00201; SO; 2.
SO SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Query Match 99.6%; Score 7291.5; DB 4; Length 1404;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1361; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAMKTLPIYLLLSVFIQVSSODLSSCAGRCGEYSRDATNCNDYNCQHMECCPDF 60
Db 1 MAMKTLPIYLLLSVFIQVSSODLSSCAGRCGEYSRDATNCNDYNCQHMECCPDF 60
QY 61 KRVCTAEISCKGRCESEFERECDDAQCCKKYDKCCPDYSEFCAEVHNPTSPSSSKAP 120
Db 61 KRVCTAEISCKGRCESEFERECDDAQCCKKYDKCCPDYSEFCAEVHNPTSPSSSKAP 120
QY 121 PPSGASQTIKSTTKRSPKPKKTKKVISEELTE----- 156
Db 121 PPSGASQTIKSTTKRSPKPKKTKKVISEELTE----- 156
QY 157 -----VDNKKNRTKKKPTPKPVVDAGSGLDNGDFKVTPTDST 197
Db 157 -----VDNKKNRTKKKPTPKPVVDAGSGLDNGDFKVTPTDST 197
QY 198 TOHKNVSTSPKITTAAPINPRPSLPNSDTSKETSLTVNKEITYETKTTTNNKOTSDG 257
Db 198 TOHKNVSTSPKITTAAPINPRPSLPNSDTSKETSLTVNKEITYETKTTTNNKOTSDG 257
QY 241 TOHKNVSTSPKITTAAPINPRPSLPNSDTSKETSLTVNKEITYETKTTTNNKOTSDG 300
Db 241 TOHKNVSTSPKITTAAPINPRPSLPNSDTSKETSLTVNKEITYETKTTTNNKOTSDG 300
QY 258 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPRAETTTKGPALTTPEKPTPTTKEPAS 317
Db 258 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPRAETTTKGPALTTPEKPTPTTKEPAS 317
QY 301 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPRAETTTKGPALTTPEKPTPTTKEPAS 360
Db 301 KEKTTSAKETOSIEKTSADLAPTSKVLAKPTPRAETTTKGPALTTPEKPTPTTKEPAS 360
QY 318 TTKKEPPTTTIKSAPTTKEPAATTTKSAPTTKEPAATTTKEPAATTTKEPAATTTKEP 377
Db 318 TTKKEPPTTTIKSAPTTKEPAATTTKSAPTTKEPAATTTKEPAATTTKEPAATTTKEP 377
QY 361 TTKKEPPTTTIKSAPTTKEPAATTTKSAPTTKEPAATTTKEPAATTTKEPAATTTKEP 420
Db 361 TTKKEPPTTTIKSAPTTKEPAATTTKSAPTTKEPAATTTKEPAATTTKEPAATTTKEP 420
QY 378 APPTTKSAPTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTK 437
Db 378 APPTTKSAPTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTK 437
QY 421 APPTTKSAPTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTK 480
Db 421 APPTTKSAPTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTK 480
QY 438 EPAPTAPEKPAATTTKEPAATTTKEPSPTTKKEPSPTTKKEPSPTTKKEPSPTTKKEP 497
Db 438 EPAPTAPEKPAATTTKEPAATTTKEPSPTTKKEPSPTTKKEPSPTTKKEPSPTTKKEP 497
QY 481 EPAPTAPEKPAATTTKEPAATTTKEPSPTTKKEPSPTTKKEPSPTTKKEPSPTTKKEP 540
Db 481 EPAPTAPEKPAATTTKEPAATTTKEPSPTTKKEPSPTTKKEPSPTTKKEPSPTTKKEP 540
QY 498 TTKSAPTTKEPSPTTKKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTKP 557
Db 498 TTKSAPTTKEPSPTTKKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTKP 557
QY 541 TTKSAPTTKEPSPTTKKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTKP 600
Db 541 TTKSAPTTKEPSPTTKKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTKP 600
QY 558 APAPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTK 617
Db 558 APAPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTK 617
QY 601 APAPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTK 660
Db 601 APAPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTK 660
QY 618 PEEDAPTTPKAADNTPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEP 677
Db 618 PEEDAPTTPKAADNTPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEP 677
QY 661 PEEDAPTTPKAADNTPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEP 720
Db 661 PEEDAPTTPKAADNTPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEP 720
QY 678 APPTPKKPAKRELAPTTTKEPTSTTSOKPAATTTKGTAPTTKPAATTTKEPAATTTKPG 737
Db 678 APPTPKKPAKRELAPTTTKEPTSTTSOKPAATTTKGTAPTTKPAATTTKEPAATTTKPG 737
QY 721 APPTPKKPAKRELAPTTTKEPTSTTSOKPAATTTKGTAPTTKPAATTTKEPAATTTKPG 780
Db 721 APPTPKKPAKRELAPTTTKEPTSTTSOKPAATTTKGTAPTTKPAATTTKEPAATTTKPG 780
QY 738 TAPPTLKEPAATTTKPAKRELAPTTTKGPTSTSDKPAATTTKETAATTTKKEPAATTKP 797
Db 738 TAPPTLKEPAATTTKPAKRELAPTTTKGPTSTSDKPAATTTKETAATTTKKEPAATTKP 797
QY 781 TAPPTLKEPAATTTKPAKRELAPTTTKGPTSTSDKPAATTTKETAATTTKKEPAATTKP 840
Db 781 TAPPTLKEPAATTTKPAKRELAPTTTKGPTSTSDKPAATTTKETAATTTKKEPAATTKP 840
QY 798 KPAATPEPTPTTSEVSTPTTKEPTTIHKSPEDESTPELSAEPPTPALENSPEKEPVPT 857
Db 798 KPAATPEPTPTTSEVSTPTTKEPTTIHKSPEDESTPELSAEPPTPALENSPEKEPVPT 857
QY 841 KPAATPEPTPTTSEVSTPTTKEPTTIHKSPEDESTPELSAEPPTPALENSPEKEPVPT 900
Db 841 KPAATPEPTPTTSEVSTPTTKEPTTIHKSPEDESTPELSAEPPTPALENSPEKEPVPT 900
QY 858 TKTPAATKPEKTTAKDKTERDLRTTPETTTAAPKMKETATTTTEKTTESKITATTTQV 917
Db 858 TKTPAATKPEKTTAKDKTERDLRTTPETTTAAPKMKETATTTTEKTTESKITATTTQV 917

Db 901 TKTPAATKPEKTTAKDKTERDLRTTPETTTAAPKMKETATTTTEKTTESKITATTTQV 960
QY 918 TSTTTQDPTTKKITTLTKTTTAPKYTTTKKITTTEINMKBEETAKPKDRATNSATTPK 977
Db 918 TSTTTQDPTTKKITTLTKTTTAPKYTTTKKITTTEINMKBEETAKPKDRATNSATTPK 977
QY 961 TSTTTQDPTTKKITTLTKTTTAPKYTTTKKITTTEINMKBEETAKPKDRATNSATTPK 1020
Db 961 TSTTTQDPTTKKITTLTKTTTAPKYTTTKKITTTEINMKBEETAKPKDRATNSATTPK 1020
QY 978 POKTPAKPKPTSTKPKTMTBVRKPKTTPTRKMTSTMPBLNPTSTRIAEMLQTTTPPN 1037
Db 978 POKTPAKPKPTSTKPKTMTBVRKPKTTPTRKMTSTMPBLNPTSTRIAEMLQTTTPPN 1037
QY 1021 POKTPAKPKPTSTKPKTMTBVRKPKTTPTRKMTSTMPBLNPTSTRIAEMLQTTTPPN 1080
Db 1021 POKTPAKPKPTSTKPKTMTBVRKPKTTPTRKMTSTMPBLNPTSTRIAEMLQTTTPPN 1080
QY 1038 QTPNSKIVEVNPKSEADAGAGETPHMLLRPHVMEVTPMDYLPRVPMOGIILNPMLS 1097
Db 1038 QTPNSKIVEVNPKSEADAGAGETPHMLLRPHVMEVTPMDYLPRVPMOGIILNPMLS 1097
QY 1081 QTPNSKIVEVNPKSEADAGAGETPHMLLRPHVMEVTPMDYLPRVPMOGIILNPMLS 1140
Db 1081 QTPNSKIVEVNPKSEADAGAGETPHMLLRPHVMEVTPMDYLPRVPMOGIILNPMLS 1140
QY 1098 DETNINCNGKPVGDLTTLNGLVAFRGHYFWMLSFSPSPARRTTEVWGISPIDYFT 1157
Db 1098 DETNINCNGKPVGDLTTLNGLVAFRGHYFWMLSFSPSPARRTTEVWGISPIDYFT 1157
QY 1141 DETNINCNGKPVGDLTTLNGLVAFRGHYFWMLSFSPSPARRTTEVWGISPIDYFT 1200
Db 1141 DETNINCNGKPVGDLTTLNGLVAFRGHYFWMLSFSPSPARRTTEVWGISPIDYFT 1200
QY 1158 RNCCEGKTFEFPKDSQYWFRTMDIKDAGYKPIFGFGGLTGOIYVAALSTAKYKMPESVY 1217
Db 1158 RNCCEGKTFEFPKDSQYWFRTMDIKDAGYKPIFGFGGLTGOIYVAALSTAKYKMPESVY 1217
QY 1201 RNCCEGKTFEFPKDSQYWFRTMDIKDAGYKPIFGFGGLTGOIYVAALSTAKYKMPESVY 1260
Db 1201 RNCCEGKTFEFPKDSQYWFRTMDIKDAGYKPIFGFGGLTGOIYVAALSTAKYKMPESVY 1260
QY 1218 FFKRGSIQOYIYKOEPOKCPGRRPALNIPYVGMQVRRRPERAIGPSQHTIRIQY 1277
Db 1218 FFKRGSIQOYIYKOEPOKCPGRRPALNIPYVGMQVRRRPERAIGPSQHTIRIQY 1277
QY 1261 FFKRGSIQOYIYKOEPOKCPGRRPALNIPYVGMQVRRRPERAIGPSQHTIRIQY 1320
Db 1261 FFKRGSIQOYIYKOEPOKCPGRRPALNIPYVGMQVRRRPERAIGPSQHTIRIQY 1320
QY 1278 SPARLAYODKGVLLHNEVSTILMRLPVNVTSAISLPIRKPDGYDYAFKDDQYNIIDV 1337
Db 1278 SPARLAYODKGVLLHNEVSTILMRLPVNVTSAISLPIRKPDGYDYAFKDDQYNIIDV 1337
QY 1331 SPARLAYODKGVLLHNEVSTILMRLPVNVTSAISLPIRKPDGYDYAFKDDQYNIIDV 1380
Db 1331 SPARLAYODKGVLLHNEVSTILMRLPVNVTSAISLPIRKPDGYDYAFKDDQYNIIDV 1380
QY 1338 PSRTARATITRSQOTLSKWNYNCP 1361
Db 1338 PSRTARATITRSQOTLSKWNYNCP 1404

RESULT 2
Q9BX49 PRELIMINARY; PRT; 1404 AA.
AC Q9BX49;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR).
GN BG174L6.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wray P.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133553; CAC36090.1; -;
SQ SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEES CRC64;

Query Match 99.4%; Score 7276.5; DB 4; Length 1404;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1358; Conservative 0; Mismatches 3; Indels 43; Gaps 1;

QY 1 MAMKTLPIYLLLSVFIQVSSODLSSCAGRCGEYSRDATNCNDYNCQHMECCPDF 60
Db 1 MAMKTLPIYLLLSVFIQVSSODLSSCAGRCGEYSRDATNCNDYNCQHMECCPDF 60
QY 61 KRVCTAEISCKGRCESEFERECDDAQCCKKYDKCCPDYSEFCAEVHNPTSPSSSKAP 120
Db 61 KRVCTAEISCKGRCESEFERECDDAQCCKKYDKCCPDYSEFCAEVHNPTSPSSSKAP 120
QY 121 PPSGASQTIKSTTKRSPKPKKTKKVISEELTE----- 156
Db 121 PPSGASQTIKSTTKRSPKPKKTKKVISEELTE----- 156

```
QY 157 -----VKDNKKNNRTKKPKPKPPVNDAGSGLDNGDFKTTPTDST 197
      |||
Db 181 KISSKSNANRELQKKLAKDNKKNNRTKKPKPPVNDAGSGLDNGDFKTTPTDST 240
QY 198 TOHNKYSTSPKITTAAPINRPSLPNSDTSKETSLTVNKETTVETKETTNNKQISTDG 257
      |||
Db 241 TOHNKYSTSPKITTAAPINRPSLPNSDTSKETSLTVNKETTVETKETTNNKQISTDG 300
QY 258 KEKTSNAKESQSTIEKTSNAKLAPTSKVLAKPTKAEETTTGPALETTTKEPTTTPKBPAS 317
      |||
Db 301 KEKTSNAKESQSTIEKTSNAKLAPTSKVLAKPTKAEETTTGPALETTTKEPTTTPKBPAS 360
QY 318 TTPEKEPTTTIKSAPTTPKBPAPTTTTSAPTTTKEBPAPTTTKEBPAPTTTKEBP 377
      |||
Db 361 TTPEKEPTTTIKSAPTTPKBPAPTTTTSAPTTTKEBPAPTTTKEBPAPTTTKEBP 420
QY 378 APPTTTSAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 437
      |||
Db 421 APPTTTSAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 480
QY 438 EPAPTTAKKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBP 497
      |||
Db 481 EPAPTTAKKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBP 540
QY 498 TTTSAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKE 557
      |||
Db 541 TTTSAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKE 600
QY 558 APPTAKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 617
      |||
Db 601 APPTAKBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTT 660
QY 618 PEBPAPTTTAAANPNTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBP 677
      |||
Db 661 PEBPAPTTTAAANPNTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBP 720
QY 678 APPTPKKPAKELAPTTTKEBPSTTSODKPAPTTTPKGAPTTTPKBPAPTTTKEBPAP 737
      |||
Db 721 APPTPKKPAKELAPTTTKEBPSTTSODKPAPTTTPKGAPTTTPKBPAPTTTKEBPAP 780
QY 738 TAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAP 797
      |||
Db 781 TAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAP 840
QY 798 KPAPTTEPTPPTTSEVSTPTTKEPTTIHKSPESTPELSAETTPKALENSPEPEVPT 857
      |||
Db 841 KPAPTTEPTPPTTSEVSTPTTKEPTTIHKSPESTPELSAETTPKALENSPEPEVPT 900
QY 858 TKTPAATKPEKTTTAKDKTTERDLRTTETTTAAPKMTKETATTTTEKTESKINATTTQV 917
      |||
Db 901 TKTPAATKPEKTTTAKDKTTERDLRTTETTTAAPKMTKETATTTTEKTESKINATTTQV 960
QY 918 TSTTTQOTTEPKITTLKTTTLAPKVTYTTTKTITTEIMNKPEETAKKRDATNSKATTPK 977
      |||
Db 961 TSTTTQOTTEPKITTLKTTTLAPKVTYTTTKTITTEIMNKPEETAKKRDATNSKATTPK 1020
QY 978 POKTTTAKPKKPTSTKKPKTPMVRKPKTTTPPKMTSMPLELNTSRIAEMLDOTTTRPN 1037
      |||
Db 1021 POKTTTAKPKKPTSTKKPKTPMVRKPKTTTPPKMTSMPLELNTSRIAEMLDOTTTRPN 1080
QY 1038 QTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVMEVTPDMOYLPRVNOGIIINPLMS 1097
      |||
Db 1081 QTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVMEVTPDMOYLPRVNOGIIINPLMS 1140
QY 1098 DETNICKGKRPVDTGLTLRLNGVLVAFRGHYFWMLSPSPSPSARITTEWGIIPSIDVFT 1157
      |||
Db 1141 DETNICKGKRPVDTGLTLRLNGVLVAFRGHYFWMLSPSPSPSARITTEWGIIPSIDVFT 1200
QY 1158 RCNCEGKTFEFPKDSQYRFTNDIDAGYKPRIFKGFGLTQIYAALSTAKYKMMPSVY 1217
      |||
Db 1201 RCNCEGKTFEFPKDSQYRFTNDIDAGYKPRIFKGFGLTQIYAALSTAKYKMMPSVY 1260
QY 1218 FFKRGSIOQYIYKQEPVQKCGRRPALNVPVGEVTOVRRRRREFRAIGPSQTHIRIQY 1277
```

```
Db 1261 FFKRGSIOQYIYKQEPVQKCGRRPALNVPVGEVTOVRRRRREFRAIGPSQTHIRIQY 1320
QY 1278 SPARLAYQDGVLAHNEKYSILMRGLPNVYTSALSLPNIRKPGGYDYAARSKOYNNIDV 1337
      |||
Db 1321 SPARLAYQDGVLAHNEKYSILMRGLPNVYTSALSLPNIRKPGGYDYAARSKOYNNIDV 1380
QY 1338 PSRTARAITRRSGOTLSKVWYNCP 1361
      |||
Db 1381 PSRTARAITRRSGOTLSKVWYNCP 1404

RESULT 3
QJUN99 PRELIMINARY; PRT; 1054 AA.
AC QJUN99;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE mRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN PRG4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikegawa S., Nakamura Y.;
RT "A novel mouse gene highly homologous to a human gene encoding
RT megakaryocyte stimulating factor precursor and cartilage superficial
RT zone protein."
RL EMBL: (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB034730; BAA92310.1; -.
DR MGD; MGI:1891344; Prg4.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR002965; P-rich_extensn.
DR Pfam; PF01033; Somatomedin_B.
DR PRINTS; PR01217; PRICHEXTENS.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR SQUENCE 1054 AA; 115991 MM; 4FC64BFA4283235 CRC64;

Query Match 52.3%; Score 3826.5; DB 11; Length 1054;
Best Local Similarity 54.9%; Pred. No. 1.1e-245;
Matches 776; Conservative 59; Mismatches 167; Indels 411; Gaps 24;

QY 1 MAKTTPLIYLLLSVIVYIOVSSODLSSCAGRCGEYSRDATNCYINCHYMECCPDF 60
      |||
Db 1 MGMTILVCSLLPLVYLIOOVSSQDSSCAGRCGEYSRDATNCYINCHYMECCPDF 60
QY 61 KRVTAELSCGRCFESFEGERECDDAOCKKYYKCCPDYESFCAEYHNPTSPSSKAP 120
      |||
Db 61 KRVCSPELSCGRCFESFEFARCECDOSCKQYKCCADHDSFCEYHNST-SSKTA 119
QY 121 PPASQOTISTTRSPKPKPKKTKVISEEITE----- 156
      |||
Db 120 TPACASDTIKSTTRSPKSP-TRTIKVESEELTEEHSDESENESSSSSSSTIRKI 178
QY 157 -----VKDNKKNNRTKKPKPKPPVNDAGSGLDNGDFKVT--TPDSTT 198
      |||
Db 179 KSSKSNANRELQKNPNVNDKNNKPNPEPAVDAGSGLDNGDFKLTLPDPDPTT 238
QY 199 OHNVYNSPKITTAAPINRPSLPNSDTSKETSLTVNKETTVETKETTNNKQISTDG 258
      |||
Db 239 PHSKVAITSPKITTAAPINRPSLPNSDTSKEASLAKETTVETKETTANNKSSA-SK 297
QY 259 EKTTSNAKESQSTIEKTSNAKLAPTSKVLAKPTKAEETTTGPALETTTKEPTTTPKEPAST 318
```

```

Db      1022 KDQYNYIDVPTARAITTRSGGTLISKIYVNCPE 1054

RESULT      4
ID      077765      PRELIMINARY:      PRT:      401 AA.
AC      077765;
DT      01-NOV-1998 (TREMBLrel .08, Created)
DT      01-NOV-1998 (TREMBLrel .08, Last sequence update)
DT      01-JUN-2001 (TREMBLrel .17, Last annotation update)
DE      SUPERFICIAL ZONE PROTEIN (FRAGMENT).
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovidae; Bovinae; Bos.
OX      NCBI_TaxID=9913;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=ARTICULAR CARTILAGE;
RA      Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,
RA      Aydelotte M.B.;
RT      "Immunodecution and Partial cDNA Sequence of the Proteoglycan,
RT      Superficial Zone Protein, Synthesized by Cells Lining Synovial
RT      Joints."
RL      J. Orthop. Res. 0:0-0(1998).
RL      EMBL; AF056218; AAD13404.1; -.
DR      InterPro: IPR000585; Hemopexin.
DR      Pfam: PF00045; hemopexin; 2.
DR      PROSITE: PS00024; HEMOPEXIN; UNKNOWN_1.
DR      SMART; SM00120; HX; 2.
FT      NON-TER      1
SQ      SEQUENCE      401 AA: 44952 MW: 86147CC9AFBB73D7 CRC64;

```

Query Match	23.4%	Score 1713.5	DB: 6	Length 401
Best Local Similarity	80.2%	Pred. No. 4.8e-106		
Matches 320	Conservative 24	Mismatches 52	Indels 3	Gaps 3
QY	964	PKDRATNSKATTPKPKPOKPTKAPKPKPTSTKPKTKMPVRKKPTTPPKRK-MTSTMPELNPT	1022	
Db	5	PKGRATNSQVTPPKPOKPTKAPKPKPTSTKPKRK-PRVKRPKTTTPPKTTTSAAMEEPTT	63	
QY	1023	SRIAEAMLOTTTRPNQTPNSKLVEYNPKSEDAAGAEGETPHMLRLPHVFMPEVTPDMYL	1082	
Db	64	S-LPEAMLOTTTRPPTPPTSEIIDVENSENEDDAAGECKPHMIFRPPVLTLPVIGTEII	122	
QY	1083	PRVPMOGLIINMISDENICMGKGVDDGLTTRNCTIVAFRCHYTWMLSPESPSPARI	1142	
Db	123	VKPSGGGGINMFESDENLGNRPVDDGLTTRNCTIVAFRCHYTWMLTTPPPPPRR	182	
QY	1143	TEVMGIPSPIDVIFRNCCEKTFEPFKDSQVYRFPNDLKADAGYPRPIFKGFGLGQIVA	1202	
Db	183	TEVMGIPSPIDVIFRNCCEKTFEPFKDSQVYRFPNDLKADAGYPRKLIKGRGGLNGKIVA	242	
QY	1203	ALSTAKYVNWPESVYEFKRGGSIQOYIYKQEPVQKCPGRPALNPVYGEMTQVRRRE	1262	
Db	243	ALSIQVYKSPRESVYEFKRGGSVOQOYTKQEPQKCTGRPALNPVYGETAQVRRRE	302	
QY	1263	RAIGSQVHTIRIQVSPARLAVQDKGVLAHNEKVSILMRGLPNVYTSALSLPNIRKPGY	1322	
Db	303	RAIGSQVHTIRIHTPRVPRVQDGGFLAHNEKVSILMRGLPNVYTSALSLPNIRKPGY	362	
QY	1323	DYAFSKDQYINIDVPSRTARATITRSGQTLKSVYNC	1361	
Db	363	DYAFSKDQYINIDVPSRTARATITRSGQTLSTNYNCP	401	
RESULT	5			
ID	Q9N4S7			
AC	Q9N4S7	PRELIMINARY;	PRT: 1079 AA.	
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)			


```
QY 165 TKKPTPKPVVDAGSGLDNGDEKVT-----TPDSTTQHNKVTSP-KITLAKPINDRP 219
DB 363 PTKSETAPBRMHYAIEIKRCOTELMYAPTEETTYAPTEETTYAPTEETTYAPTEETP 422
QY 220 SLPPNSDTSKETSITVANKETTVETKETT---TNKQSTDGKEKTSAKETSIKTSK 276
DB 423 YEP-----TEETTYTPTEETTYAPTEETTYAPTEKTYAPTEETTYAPTEETPYEPT 477
QY 277 DLAPTSLVLAKPPTKATETTKGPAITPKKEPTPTPKKEPASTPKKEPTPTIKAPPTPK 336
DB 478 TYAPTKETTYAP---ETTYASTTEETTYAPTEETTYAPAEETPYEPTET---TYAPTEET 533
QY 337 EPAPT--TKKSAPPTPKKEPAPT--TTKEPAPTTPKEPAPTTPKEPAPTTPKSAPT----- 387
DB 534 TYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTETTYAPTEET---YAPTEETMYA 592
QY 388 ----TPKEPAPTTPKKAPPTPKKEPAPTTPKEPTPTPKKEP-----AP--- 426
DB 593 PIEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYASTTEETTYAPTEE 652
QY 427 TTKEPAPTTPKEP-----APTAKKAPPTPKKEPAPTTPKEPAPT--TTKEPSPTPK 478
DB 653 TYAPAEETPYEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYE 712
QY 479 P-----APT--TKKSAPT-----TTKEP-----APT--TKKSAPT--TPKEPSPT 512
DB 713 PTEETTYAPTEETTYAPTEETMYAPIETTYGPTTEETTYAPTEKTYAPTEETTYAPTEE 772
QY 513 TTKEP-----APTTPKEPAPTTPKKAPPTPKKEPAPT--TPKEPAPTTPKKAPPTAPKE 564
DB 773 TTYEPTGTEETTYAPTEETTYAPTEETTYAPAEETTYAPTEETPYEPTETTYAPTEETPYE 832
QY 565 PAPPTPEAPPTPKKILPTPEKLAPTPEKAPPTPEELAPT--PPEEPTPTPEEP- 621
DB 833 PTEET--TYPTTEETTYAPTEETTYAPTEKTYAPTEETTYAPAEETPYEPTETTYAPT 890
QY 622 ----APT-----TPKAAPNTPKKEPAPTTPKEP-----APTTPKEPAPTTPKE 660
DB 891 KETTYAPTEETTYASTTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEET 950
QY 661 TAPTPKGIAPT--TLKEPAPTTPKKAPK-ELAPT--TTKEPSTTSKAPPTPKGTA 715
DB 951 YAPTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYAPTEETMYAPIETTYAPT 1010
QY 716 PTPPEAPPTPKPEP-----APTPKGIAPT-----TTKEPAPTTP 751
DB 1011 EETTYAPAEETPYEPTETTYAPTEETTYAPTEETTYASTTEETTYAPTEETTYAPAEETP 1070
QY 752 KKP-----APKE---LAPT--TTKGPSTTSKAPPTPKETAPPTPKKEPAPTTPKKA 800
DB 1071 YEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTETTYAPT 1130
QY 801 PTPPEPTPTSEVSTPTTKKEP--TIHKSDESPELSAPTPKA-LENSEKPEG--- 854
DB 1131 EETTYAPTEETMYAPIETTYGPTTEETTY-APTEATTYAPTEETPYAPTEETTYEPTGET 1189
QY 855 --VPTTKPATRKEMTTAKDKTERDLRTP---ETTTAKMKMKATATEKTES 908
DB 1190 TYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTETTYAP---TEETTYEPTEE 1244
QY 909 KITATTTQVSTTQDPT-----PKITT---LKTTLAPKVTTKKITTTEIM 955
DB 1245 TYAPTEETTYAPTEETTYAPTEETMYAPIDETITYGTEETTYAPTEATTYAPTEETPYA 1304
QY 956 NKDEETAKPRDRAI--NSKATTPKP-OKPTKAPKKPT-----STKRP-KTMPR 999
DB 1305 PTEETTYEPTGETTYAPTEETTYAPTEETTYAPAEETPYEPTETTYAPTEETTYEPT 1364
QY 1000 VRKPKITPTRKAKTSTMPELNPTSRIAEAMIQITTR--NOTPMSKIYEVNPKSSEDG 1055
DB 1365 TDEPTDEPTDE--PSDEPTDEPTDEPTDLPTDEPTPCDNOGINGIVENKVRYNNG 1420
```

```
RESULT 9
Q9SPMO ID Q9SPMO PRELIMINARY; PRT; 1315 AA.
AC Q9SPMO;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DE 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EXTENSIN-LIKE PROTEIN.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACO clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=POLLEN;
RA Stratford S., Barnes W., Golubiewski A., Colter R., McCormick S.,
RA Hohorst D., Gao M., Showalter A., Bedinger P.A.;
RT "Pollen extensin-like (Pex) Genes in a Monocot and a Dicot.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159297; AAD55980.1; -.
DR Interpro; IPR001611; LRR.
DR Interpro; IPR001998; Xylose_isom.
DR Interpro; IPR002965; P_rich_extensn.
DR Interpro; IPR003592; LRR_out.
DR Pfam; PF00560; LRR_3.
DR PRINTS; PR01217; PRICHEXTENS.
DR PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR SMART; SM00370; LRR; 3.
SQ SEQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;

Query Match 11.5%; Score 839.5; DB 10; Length 1315;
Best Local Similarity 24.4%; Pred. No. 1,3e-47;
Matches 265; Conservative 101; Mismatches 491; Indels 227; Gaps 34;

QY 73 RCFSEFREGREDCDADCKKRYDKCCPYESFCVHNHPTSPSSK-KAPPSGASQTIKS 131
DB 390 QCAVILARPEVDCS-----KHVCAQYPTPGGPPSSVYGRPSVYGR----- 432
QY 132 TTKRSPKPPNKKTKKVIIESEITEVKNKKNRTKKRPPKPVVDAGSGLDNGDEKVT 191
DB 433 --PAAPAMPPTPHPPDVSEPIPEPSVPAPAPARMPTLRSPADE-----YIP 480
QY 192 TPDSSTQHNKVTSSKTIITAKPILNPPSLPNSNDTSKESLIYANKETIYETKETTNNK 251
DB 481 TTPVPA--KSPGTSPPASRGAP--PLQADPPPAASSPPAPV----- 518
QY 252 QTSIDGKEKETSIAKETQSIKTSKADLAPTSGLVLAKPTEKPAETTKGPAITTPKEPTPT 311
DB 519 -----KSSPPPAVILPAPKATSPSPAPV-ASPPPEAPVSSPPQVKSPPPPAPVA 568
QY 312 PKPEASTPKPEPTPTTKSAPPTPKKEPAPTTPKSAPPTP-KEPAP-----TTTKEPAP 363
DB 569 SPPPKMSP--PPPARVASPPPLKMSPPPAVVASPPQPLKSPPVILMISTSVKSP 626
QY 364 TTKEPAPTTPKEPAPTTPKSAPPTPK-----PAP---TTP-----KKAPPTPKKEPAPTTP 413
DB 627 PVPAVSPPPVKSPPPLAPVSSSPVKKLPLPAAGSSTPPEBEKPTPTPKVSSPPE 666
QY 414 KE--PTPTPKKEPAPTTPKEPAPTTPKKEPAPPAKKAPPTTPKEPAPTTPKKEPAPTTPKEPS 472
DB 687 KSLDPTLTSSPPOEKPTPSTPSKPPSPVETLPPKSSSPPEPVSSPPQAPKSSS 746
QY 473 PTPPKKEPAPTTPKSAPPTTPKEPAPTTPKSAPPTPKKEPSPTTPKEPAPTTPKKEPAPTTPPK 532
DB 747 P-----PAPVS--SPPPLKSSPPVPESSPPPTPKSSPLAPVSSPPQVKEKTSPPAPVS 799
QY 533 PAPTPKEPAPTTPKEPAPTTPTKKAPAPTAKEPAPTTPKETAPTTPKKLAPTTPKEKLAAPT 592
DB 800 SPPPTPKSSPPLAPVSSPPOVERKTSPPAPVSSPPPTPKSSPPLAPVSSPPQVKEKTSPP 859
```

```
QY 593 TPKPAPPTPEELAPPTPEEPPTPTTPEEPAPPTPKAAAPPTPKAPPTTKEPAPTTPKE 652
DB 860 APVSSPPELEKPPSSPSSVSPPTTVKSSPPAPLSSPMTPKSSSPAPVSSPPEAKS 919
QY 653 PAPTTEKAPPTPKGA-----PTTKEPAPTTPKAPKAPLAPTTTKEPTST 701
DB 920 SPLAPLSSPSPKSSPSSPSPVWPKTSPAPATVSSPPTPKSSPP--APVSSPPVVK 976
QY 702 TSDPAPPT-----TPKGTAPPTP-----KEPAPTTPKEPAPTTPK--GTAAPTTPKE 746
DB 977 SSPAPVSSPPTPKPLPPAPVSSPSPVVKSSPPTPVSSPPTPKPLPPPTPVSSP 1036
QY 747 APPTPKPAPKAPLAPTTKCPSTTSOKPAPT--PKETAPTPK--EPAPTTPKPA 800
DB 1037 PTPXKPLPP--APVSSPPVVKSSPPAPVSLPPTPKPKPPRTVSSPPVVKCCPP 1093
QY 801 PTPPTPTPTSEVSTTTTKEPTTIHKSDESPDELAPPTKALENSKPEGVPTTK 860
DB 1094 PTLVSSPSPAPKSLPPTPVSSP-----PPEVKSPPPTPVSSPAPKSSPPT 1143
QY 861 PAATKPEMTTAKDKTTERDLRTTP-----ETTTAPKMKETATTEKTESKTATTT 915
DB 1144 PVSSPP-----ELKSSPPAPVSSSPSAPKSSPPAPVSLPPEVKSSPPA 1191
QY 916 QVSTSTTQDTPPKITTLKTTTAPKVTTKTTTTEINKKPEETAKPKORATNSKAT 975
DB 1192 PISPPPPAKSP-----PPA-----PKMSLPVVKSSPPAPV 1225
QY 976 PKQPKPKAKKPTSTKKPTMRVKKPTTPPKMTSTMPELNPTSLAEALQTTTR 1035
DB 1226 SSPPPKSPSPAPVSSPPAP--VKPPLPPAPVSSPPAPV-----TSA 1270
QY 1036 PNOTPNSKLVEVNPKSEDAAGAEETPHMLRPHVFEVTPDMDYLPVRYNQIILNPM 1095
DB 1271 P-----PKKEEDSTA-----PPAEALPPSPENDIILPPI 1299
QY 1096 LSDE 1099
DB 1300 MANK 1303

RESULT 10
Q20007 PRELIMINARY; PRT; 1274 AA.
AC Q20007;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (T-EMBLrel. 08, Last annotation update)
DE COSMID F35A5.
GN Caenorhabditis elegans.
OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdillida; Rhabdilloidea;
OC Rhabdillidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA MEDLINE=94150718; Pubmed=7906398;
RA Wilson R., Alnecough R., Anderson K., Baynes C., Betks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Garder A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shonkeen R.,
RA Smailton N., Smith A., Sonhammer E., Staden A., Sulston J.,
RA Thelery-Mleg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans."
RN Nature 368:32-38(1994).
RP [2]
SEQUENCE FROM N.A.
```

```
RC STRAIN-BRISTOL N2;
RA Leimbach D.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U46675; AAB52641.1; -.
SQ SEQUENCE 1274 AA; 138065 MW; 07F6B0D292C1799F CAC64;
```

Query Match 11.4%; Score 835; DB 5; Length 1274;
Best Local Similarity 27.5%; Pred. No. 2.8e-47;
Matches 310; Conservative 111; Mismatches 446; Indels 260; Gaps 60;

```
QY 95 KCCPDYSEFCAEVHNPTSPSSKKAP-----PPSGASQTIKSTTKR 135
DB 189 KAAPSKS-----HDPVLPPTPKNPAPKPKPWEDEVPEETKEPEATRKVPALKKK 242
QY 136 SP-----KP-PNKKKKKVISEETIEVKNKKNKTKK--KP-----TP-----KPPVDEA 179
DB 243 EPSTSVKPVGSDPSPTKKVPAKKEPEVPTPIKNTKKMKPWEDEVPEEVEKKEPVEKK 302
QY 180 GSGIDGND---EKVTPDSTTOHKNVSTSPKITTAKPI--NPPSLPPNSDTSKETS 234
DB 303 APVLKKKDPAPAAKARDPSPSKAPKKVESSPVVPTPVKNPVKKYKPPWEVDEPAEE 362
QY 235 VNKETIVETETTTNKTSTGDKETKTSKENSISIKTSKDLAPRSKVLAKPTPAE- 293
DB 363 VKKPSAPEKTPVLKREPESSSTPSSDSPPKAAAPVAKPPOSSPKATPQADPAQE 422
QY 294 ---TTTGPA-----LTTPKEPTPTTP-----KEPATTPKEP-----TP 325
DB 423 VPPTPVKNPVKKYKPPWEVDEDEVPEVQDEAPAKKTTPVLKREKPAKAKTATSKTP 482
QY 326 -TTKSAPTTPKEPAPPTTTSAPPTTKEPAPTTPKEA-----PTT 365
DB 483 ETPEKKDPVVRPOSSPKVAAKPDQAQPA--TPVKNPVKKWRPWEDEVPEADSVKPTD 541
QY 366 PKEPAPTTPKEAPT-----TTKSAPTTPKEP-----APTTPKAPPTPKP-----PA 409
DB 542 AKKTPSLAKKDPAPAKSLKPKADTKAPAKERDPSPKVAPTAPEKTPVLAKKEPAGPA 601
QY 410 PTPPEPTPTTPKEPAPTTPKEPAPTTPK--EPAPAPKAPAP-----TPKEPAPTTPKE 462
DB 602 DSKTEPEKSKRPDPSKKAVPAKPVKTEVAPAAVKKPEPISKPKDTAPAKAPNSPV 661
QY 463 PAPTTPKEP-----SPTTPKEPAPTTPKSAPTTTKEPAPTTPK--KEPSTTPKEP 517
DB 662 P-PTPVKNPVKKWRPWEVDDAPAKPVSLPEPEKK--TPVLAKKAPTKPDSEAAADPVSGP 719
QY 518 APPTPK--EPAPTTPKAPAP--TPKEPAPTTPKEPAPTTPKAPAP--PREAPAPTTPKE 572
DB 720 SSKDPKLAKRAVAKRDPSPKAVPIKAPPT--EVPAVKKKEPVAKSHDPSPKAP- 776
QY 573 TAPTPKPLPPTTP-----EKLAPTPEKAPPTTPEELAPPTPEEP-----TPPTPEEPA 622
DB 777 AEPSP--VVPPTPVKNPVKKWRPWEVDDAPAPAPVNVPEEKKTTPVLAKTTPVAPKPDPS 834
QY 623 PTPPKAAPNTPKEPAPPTTPKEPAPT--PKEPAPTTPKETAPPTTKGAPTTPKAPPTT 681
DB 835 PKKAVPAKPSKTADAPVSVKPKPEVSKPEPSKKAKEPSPVVP-----TPPVKNPVKK 890
QY 682 PKKAPKAPLAPPT--TTKEPTSTSDKAPPTPKGAPPTTPKEPAPTTPKEPAPTTPKGTAP 740
DB 891 -KPPWEVDDDETEVKKPSE--PEKKTTPVLAK--KEPEKPD-APKVAAPKPDPSPKAPV 945
```

```

QY 741 TTLEKAPAT-----TPKKAP-----KELAPTTTGGPTS-----TTSUKP-----AP 777
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 946 E-KEPAKVAKAPRDLSPKALIPANTQEARPTPVKMKPVMEDDEDEAPESAP 1003
QY 778 TTKET-----APTTKEKAPATTPPKAPPTTPTTSEVSTPTTKEPTTIHNSP- 830
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 1004 EPEKKTIVLAKKAKAPKRPD-----SPKKAAPVAAR-PDKPIEIV-PPPVKKNPVKMKPWP 1058
QY 831 ---DESTPELSA-EP-----TPKALENSPREPGV-----PTTKPPAAT----- 864
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 1059 EDDDEBPESVAPPEKKTIVLAKKAPATKPKATKDESEADAPVSGPTSKDKLSKKAPE 1118
QY 865 KPEMTTAKDKTTERDLRTTPEET--TAAPKMTKETATTTETTESKITATTTQVSTTT 922
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 1119 KKPRTTDPKDKLKPSAPKKEKAPDEPAAPKMKPVMDDEDEADPTVAPSKKPDTE 1178
QY 923 QDTTPEKTTILKTTLAPKVTYTTKTTTTEIIMKPEETAKPKORATNSKATTKPKOKPT 982
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 1179 DPADPLG-----GPKTKDPK-----LNKKAPEKPTKE-----BKPKESV 1213
QY 983 KAPKKTSTKKKTMPTVVRKPK-----TTTPRKMTSTMP 1017
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 1214 KEPPKTEPPKP-AAKMKKRPMEDEDEPADTTMPAKKPTEDP 1259

RESULT 11
QY 09LIE8 PRELIMINARY; PRT; 1480 AA.
AC 09LIE8;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE SMILABILITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxID=3702;
RN 11
RC SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA.
RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 121
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA.
RX PubMed-10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety pL
RT TAG and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL; AF001306; BAB03062.1;
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 1480 AA; 147153 MW; DIAOC679F155E732 CRC64;

Query Match 11.3%; Score 830; DB 10; Length 1480;
Best Local Similarity 29.0%; Pred. No. 7e-47;
Matches 321; Conservative 62; Mismatches 539; Indels 186; Gaps 48;
QY 110 PTSPSSKKAAPRGASQTIKSTKR---SPKPKKKTKKVISESEITTEKDKKKRKT 166
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 46 PRKPRAPKRNPRK--PRTIKRPPKRNPRKPTVAKRP--HKRPTKRNPRKPTK 100
QY 167 KKPTRPVUDEAGSGDNGDFKVTPTDSTQHNKVSTSEKITTAAPINRPSLP---- 222
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 101 RNHRKRPRT-----KRRNHRKRNHRKRNHRNVR--HNHRKRPRTKH 140
QY 223 -----PNSDTSKETSLTVNKETVEIKETTTNKQISTDKKKTSAKETOI 270
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 141 RNHRKRNHRKRPRTIKRPPKRPVSVAKRPSPSTPKRPTTNPRSTQRPRTNRRPSPTPR 200

```

```

QY 271 EKTSAKDLAPTISKV--LAKPTKAEITTKGPAITTKPEPT--TTP--KEPASTTKPEPT 324
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 201 PVASPMATPTTQMPRIATP-PIAKSPVATPIATPTPIATPIATPIATPIATPIATPI 259
QY 325 PTTIKSAPTTKEKAPATTT-----KSAPTTKEKAPATTT-----KEPAPT 365
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 260 PPIIMPRIATPPVAAPPTINPDISKPVATP-----PTTPIAKPIATPIISIPVATPT 315
QY 366 PKEAPTTTKEPA-----PTTKSAPTT-KEPAPTTKPKAP--TTPKEAPATTKE 415
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 316 PAATPITITLPPAKPVVAISPIVT--PVTIAQPVATPTATPPVATPIATPIATPI 373
QY 416 PPTTPKAPATTTKEKAPATTP-KEBAPAPKAPATTTKEKAPATTTKEKAPATTT--KE 470
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 374 PISTPISSEVATPTATSPITKTPPAKPVATPIATPIATPIATPIATPIATPIATPI 433
QY 471 PSPTPKEKAPATTTKSAPTTKEKAPATTTKSAPTTKPEP--SPTTTKEKAPATTTKEKAP 527
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 434 PVATPPTTTPPTAT--PVAKPVETPPIATPTAKPPISTPDISKPVATPTPAATPPI 490
QY 528 --TTPKAPATTTKEKAPATTTKEP--ATTTKKAPVAP--KEBAPTTKETAATP----- 577
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 491 TTPTPVAKPVATPIATPIAPVAKPVVATPTATPIATPIATPIATPIATPIATPIATPI 550
QY 578 --PKKLP--TTPKLAATTEKAPATTTPEELAPTT--PEEPTPTTKEEAPATTTKKAAPN 632
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 551 AKPVVATPTTTPATPIAPVAKPVVATPTATPIATPIATPIATPIATPIATPIATPI 610
QY 633 T--PKKAPTTKEKAPATTTKEKAPATTTKEKAP--TTPKSTAPTT--LKEAPATTPK 684
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 611 TTPPAKPVATPIATPIATPIATPIATPIATPIATPIATPIATPIATPIATPIATPI 670
QY 685 PAPKELAPTTKEKAPATTTSDKAPATTTKGTAP--TTPKEKAPATTTKEKAPATTTKGTATTL 743
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 671 AIPPIATPIAPVAKPVAT--PPTATPIATPIATPIATPIATPIATPIATPIATPIATPI 727
QY 744 KEPAATPKKAPKELAPTTTTCGPTST--TSDKAPATTPKETAATP--PKKEBAPTTPKKA 800
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 728 --PPTATPIAPVAMPPIATPIATPIATPIATPIATPIATPIATPIATPIATPIATPI 785
QY 801 PTPP-----ETPPTTSEVSTPTTKEPTTIH--KSPDESTPELSAEPPTKALENSPKE 852
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 786 ATPPTATPPVATPIAKPVATPTTATPTATPVAKPVATPIATPIATPIATPIATPI 845
QY 853 PGVPTTKTPATKPEMT--TTAKKKTTERDLRTTETTTAAKMKTKERATTEKTEKSTIT 911
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 846 SKPVATPPAPATPTTTPPAKPVVATPIATPP--IAKPVATPPATPIATPIATPIATPI 903
QY 912 ATTQVSTTTQDTTPEKITTLLKTTTLAPKVTTTKKT-----ITTEIIMNKPEETAKPRDR 967
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 904 PVATPIATPIATPIATPIATPIATPIATPIATPIATPIATPIATPIATPIATPIATPI 960
QY 968 ATNSKATTPKOKPTKAPKPKTSTKKKPTMVRVAKPKTPTPRKMTSTMPELN----- 1021
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 961 ATSPVATPIAKPPTTTP--PTAT--PVVAMPPIATP--PRAKPVATPIATPIANPEVEK 1014
QY 1022 --TSRIAE-----AMQTTTTRNOTPNSKL--VEYNKPESEADGAGETPIIMLRPH 1069
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 1015 VATPIAKPPTVLRPIAKPVETSPATPTATPPVATPP-----VVKRP 1059
QY 1070 VFMEPVATPMDYLPRVNOGIIINPMLS 1097
| 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 | 1111 |
Db 1060 VATPITTKRPVATPVATINPTAMPPIYV 1087

RESULT 12
QY P70670 PRELIMINARY; PRT; 2187 AA.
AC P70670;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)

```

DE NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
DE MUSCLE-SPECIFIC FORM GP220).
GN NACA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96312450; PubMed=8698236;
RA Kotov W.V., St-Arnaud R.;
RT "Differential splicing in of a proline-rich exon converts alphaNAC
RT into a muscle-specific transcription factor".
RL Genes Dev. 10:1763-1772(1996).
DR EMBL; U48364; AAB18734.1; -
DR EMBL; U48363; AAB18732.1; -
DR MCD; MGT:106095; Naca
DR InterPro: IPR002715; NAC.
DR InterPro: IPR003037; TS-N.
DR Pfam: PF01849; NAC; 1.
DR Pfam: PF02094; TS-N; 1.
SQ SEQUENCE 2187 AA; 220599 MW; 003646A864DEBFD CRC64;

Query Match 11.3%; Score 828.5; DB 11; Length 2187;
Best Local Similarity 28.4%; Pred. No. 1.3e-46;
Matches 313; Conservative 126; Mismatches 436; Indels 229; Gaps 52;

QY 109 NPSPSSSKAAPPS-GASTIKS-----TTKRSKP--PNKKTKKYESEIT 155
DB 746 SPT-PPSSKAGPVSTGAPSPGAPIVPESSISSQVPAEILPQKTEVTASLIS 804
QY 156 EVDNKKNRKTKKPPVVDGSGLDNGDFKVTPTDSTQHNK-----VSTSPKIT 210
DB 805 AVGSPKVDPLMSDVTPTSPKKTSA-----TAVPKDSATLSLKSVAVYSLSPKA 855
QY 211 TAKPINDRPSLPNSDTSKTSKSL--TVNKET-----TVETKETTNNKOTSDGKEKT 262
DB 856 PVPSPNATVTPTEIPTSLKNALAAATPKETLATSIRKVTSPSPQKPKSVSLGAPAMT 915
QY 263 SAKETOSIKTSKADLAPEKVLAKPPKPAETTKGPALETPKPTTKEPASTPKPE 322
DB 916 SKKAT---ELAAKSDVSPSO--PEKEYPLQHV---PPTSPKSPVSDTSLGALTSPPK 967
QY 323 PRETTIKSAPTPKE--PAPTTKSAPTPK-----EPAPTTKEPAATTPKEPAT 372
DB 968 GPATLAEPTLVKPKSKPAASKKTPATPSPGVTAVLELIPCKSKAPATAKESAT 1027
QY 373 TTKEPAPT--TKSAPT-----TKREPAATPKK--PAPTTKEP 408
DB 1028 SSSKRAKPTAVSKKEIPSKGVTAVLEISLPLKETSATGGEKSSSPKRSPTAGPKE- 1086
QY 409 APPTPKPEPTTTPKEPATTPKEPATTPKPAATPKKPAAPT--PKEPAP--TTPKE 462
DB 1087 --TPGCVTVAPPEISLPKETPOMATPNSLAASSQKRSKTSVPEPTPGGVTAMPLE 1144
QY 463 PAPTTKEPSPTPK-----EPAPTTKSAPTTTKEPATTT--KSAPTTPK-----E 508
DB 1145 IPSAPQAKAPATAVKOIPTEDEAVTILAGSPLSPKKASKAAPAKREARATSVGVIAVSGE 1204
QY 509 PSPTTKEPATTPKKEPATTPKPKAPAPT--PKE-----PAPTTPKEPAP-- 551
DB 1205 ISSPKKTSKTAAPKENSATLGPERSKPTAPKETPATSEGVYAVPSEISPPPTASK 1264
QY 552 -----TTTKKAPAPAKREPATTPKEPATTPKKTLPPTPEKLAATTPKEK--PAPTPELA 606
DB 1265 GVPTTLTPKAPNALAE--SPASPKVPTKTAPEETSTTP-----SPQKTPKVAAGPREAS 1317
QY 607 PTPPEPTPTT-----DEEPATTPKAAAPNTPKEPATTPKE--PAPTTPKEPAT 657
DB 1318 ATPSKKTPKTAAPKENSATSEGVYAVLEIPSPKAPATAKKEPPAS--PEGATTA 1375
QY 658 PKEPATTPKGTAPTTLKEPATTP-----KKPAPKELAPTTTKE 697

DB 1376 PVOIPSPRGRSKKAGSKE--TPTPSPGVTAAVLEIPSSKTSKMASPKETLVTPSSK 1434
QY 698 PTS-----TTSKRAPATTPKGTAPTPKREPAPTPKE--PAPTTPKKTAPTTLKEPAT 750
DB 1435 KLSQTVGPKETSLEGATVAVLEIPSSHKKAPKTYDPQVPLTPSPK--DAPTTLAE--SPSS 1492
QY 751 PKRPAPKELAPTTKGPSTTSKAPATTPKETAAPTPKKEPATTPKPAATTPPEP--- 807
DB 1493 PKK-APRTAAPPSER--VTVPPEKPA--TPKASGTSKVPVPAEIOEVAVSSSEPTVP 1549
QY 808 --PPTSEVSPPTTK-----EPTTHKSPDESTPLSAEPPKALENSPKE--PGVPTT 858
DB 1550 AVPPVKNPSSHKKTSKTELEKAPATLPSPSPKIPSSKKAPRT--SAPKEPPASPSI 1607
QY 859 KTPAATPEMTTIAKODTTERDLPETTTAAPKMKKEATTEKTEKITATTTQVT 918
DB 1608 K-PVTT--SLAQTPAPSLQAPSTTIPKENLAAPV---LPVSSKSPAPAPARSASLSP 1660
QY 919 STTTQDTPPKITTLTKTTLAPKVTTTKTITTEIMKKEEPAPKPDRAATNSKATTPK 978
DB 1661 ATAPQTAPEKATTPISCKKAATETPIETSPASLEGAKETSE-----TSVKVLMSSP 1716
QY 979 QKPTAKPKPTSTKPKPTMPRVKPKTTPPKMTSTMPELNPTSRIAEMLQTTTPNQ 1038
DB 1717 -----PKKASSSKRASTLP-----ATTLPLSKEASVLS-----PTA 1747
QY 1039 TPNSKLVEVNPKSDAGAGETP 1062
DB 1748 TSSGKDSHSPVS-DACSTGTTTP 1770

RESULT 13
041805
ID 041805 PRELIMINARY; PRT; 1188 AA.
AC 041805;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE EXTENSIN-LIKE PROTEIN PRECURSOR.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=873; TISSUE=POLLEN;
RA Rudinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z34465; CAA84230.1; -
DR Mendel; 14346; Zeama; 2368; 14346.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR Pfam: PF00350; LRR; 3.
DR SMART; SM00370; LRR; 4.
KW SIGNAL.
FT SIGNAL. 1
SQ SEQUENCE 1188 AA; 120980 MW; 2C77C7F8D7130149 CRC64;

Query Match 10.9%; Score 800; DB 10; Length 1188;
Best Local Similarity 28.7%; Pred. No. 5.4e-45;
Matches 246; Conservative 64; Mismatches 401; Indels 146; Gaps 36;

QY 82 RECCDQCKKYDKCCDYSEFCAEVINPISPPSSKAPPSGASQITKSTTKSPRPN 141
DB 397 RPVDCS-----KHVCAGYPTPGGPPSSVPKGAASAPMPS-----PHTPPD 439
QY 142 KKKTKKVTSEPEVKNKKRNTKPKRPVVDGSGLDNGDKVTTPTDSTQHN 201
DB 440 -----VSPPEPLPEPSPVAPAPMPMTPIHSPAD-----DVVPTPVPVGGKSP 483


```

QY 374 -----TKRP---APT-----TKSADPTP 389
DB 1291 EGFMADEPNNSKEYRCVRRNKGFTSIPFOCGAGTWDODLOTGNHNNNCSGTGSTSTP 1350
QY 390 KEPAATPKKAPATTPKEPAATTPKEPTPTTPKEPAATTPKEPAATTPKEPAATPKAP 449
DB 1351 KPPC--EPATNGTATSTSTTTP--PTTTDLPTSTGLP--PTTTELP----- 1398
QY 450 TTPKEPAATTPKKEPAATTPKEPSPTTPKEPAATTPKSAPTTPKEPAATTPKSAPTPK-- 507
DB 1399 TTTDLPTTTTTLPLPTTTTSLPTTTTGLPTTTTGQPTTTLSSTETSTYTTSST 1458
QY 508 -ESPPTTKKAP----- 523
DB 1459 TOPSPTTTPMLPAGTCTEGYADPEDCRKYRCINAGASYRKNYFTCPKGTGNNEV 1518
QY 524 -----EPAP---TTPKKAPATTP---KEPAATTPKE--PAPTTTKKAPATAKEPAATTP 569
DB 1519 QTCDYVENITRCKSLPAPETTTTPSEESKDPGSTTQSTDEPTTVTKPI--TKPTE--EPST 1576
QY 570 PKETAATTPKLLPTTPKEKAPATTPKEKAPATTPBELAPATTPPEPTPTTPPEPAATTPKAA 629
DB 1577 EKQOKPTTQVPEKPTTEE--PEKQOKPT--TTEYQOKPTTEETPEKQOKPT--TTEYQ 1632
QY 630 APATPKKEPAATTPKEPAATTP--KEPAATTPKETA---PTPK--GTADTTKEPAATTP 661
DB 1633 KPTTEETPEKQOKPTTTEYQOKPTTEETPTTISQYNTPTTSVQYNTPTTPIVET 1692
QY 682 PKPAPKELATTPKEPTSTSDKAPATTPKGAATTPKAPATTPKE--PAPTTPKATAP 740
DB 1693 TSPGK--PTTGTETTTT--LPSTTDAIOEPTTKKPEPTTTESESTPESST 1748
QY 741 TTLKEPAATTPK--PARKELA-----PTTKGPTSTSDKAPATTPKETAATTPK 769
DB 1749 TLDEPQPNNSSEGFPEPDECSRYRCYDAKNGKYVAFKCKGKGVMTSTETCN 1808
QY 790 EPAPTPKKEPAATTPETPTTSEVSTPTTKPTTHKSPTDESLAEPKALENS 849
DB 1809 YADQVGN-----CSSGQTTPTTTEPTTSTESTSSGK---ETTSKAPEN- 1852
QY 850 PKEPVATTPATPKEMTTAKKTEEDLTTPETTTAAKMKKETATTEKTESK 909
DB 1853 -----TTTA--PETTT-----SSEPTTT---VASETTTTSGT--- 1883
QY 910 TTAATTTQVSTTQDTTPEKITTLKTTTLAKVYTTTKITTEIMNKPEETAKPKDRAT 969
DB 1884 ---TTTATPETTTKPKP-----ETTTTGETSTSKSPTTE----- 1918
QY 970 NSKATTPKOKPTKAPKPKPTSTKPKTMP 998
DB 1919 -----SPAPSTNTAP-----CPETGP 1935

RESULT 15
O9XDH2 PRELIMINARY: PRT: 763 AA.
AC O9XDH2:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Esplita C., Lacleite J.L., Mondragon-Palomino M., Martens A.,
RT Zhang Y., Moreno C., Singh M.;
RT Cloning and characterization of a new member of the PGRS family that
RT is a useful marker of polymorphism in Mycobacterium tuberculosis."

```

```

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF071081; AAD41594.1; -.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR002965; P-rich_extensin.
DR InterPro; IPR003882; Pfic1l_extensin.
DR PRINTS; PRO1222; ATROPHIN.
DR PRINTS; PRO1217; PRICHEXTENSIN.
DR PRINTS; PRO1218; PSTLEXTENSIN.
DR SEQUENCE 763 AA; 75034 MW; 39168EC45A5916F8 CRC64;

Query Match 9.8%; Score 715; DB 2; Length 763;
Best Local Similarity 28.2%; Pred. No. 1.5e-39;
Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

QY 288 PTPKAETTTGPLTTTPKEPTPTTPKEPAATTPKEPTPTTKASAPTPKKEPAATTPK 347
DB 3 PVP-----APRALPLPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 46
QY 348 TTPKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKE 407
DB 47 PCF--PAP-----PAPPKSKAPFPVVPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 90
QY 408 PAPTPKEPTPTT---PKEPATTPKEPAATTPK--EPAPATPK--KEAPTPKKEPAATTPK 461
DB 91 SRPALPCPPPPVVIDPDEPAPAPVPAPAPNSPPFPFPAPKFPVAPVP--PVENSP 148
QY 462 EP--APTTPKEPTPTTPKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTP 519
DB 149 FPPPPAPALNPAP---PAPPLANSPLPAPAPAPAG--PAPAPWVPVAPAPSKPA 201
QY 520 TTPKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKEPAATTPKKE 575
DB 202 SPAPAPAP-----PMATPMEFPRLPVPVPDPISKETRRAPAPAPAPAPAPAPAPAPAP 256
QY 576 TTPKLLPTTPKEKAP-----TTPKEPAATTPBELAPATTPPEPTPTTPPEPAATTP-- 625
DB 257 PVPNKKTPAPAP--APVAVAAVLVACPLPLPNNHNPAPAPAPAPVGVPLAPLPSH 313
QY 626 ---PKAAPATTPKEPAATTP----- 639
DB 314 PAPSPAPVEGPLAPLPLDISCRPVSWKGSFTTSTGCRVCSSEVLGALNPSRPSRL 373
QY 640 --TTPKEPAATTPKKEPAATTPKETA---PTTKGTA---PTTLKEPAATTP--KKAPKE 689
DB 374 TTTTALPALPAPLPLPPLPPLINTAVPRLPLRVNLAAPLPLAPLPLPISGVAPAP-- 431
QY 690 LAPTTPKEPTSTSDKAPATTPKGTAPTPKEPA-----PTPKKEPAATTPKGTAPTL 743
DB 432 --PIPPKPTWTPPLAPAPPEPK--TVPLPFGSCPESEKPNPAPPEPEPKSAPLAP 488
QY 744 KEPAATTP---KKAPKELAPATTTKGP--TSTSDKAPATTPKKT---APTTPKEPAATTP 796
DB 489 APAPSPMSAVKVPSPPIPPAPAPAPASMPALPAPAPSPATRLCPPLPSPAPNSP 548
QY 797 KKAPATTPETPTTSEVSTPTTKKEPTTHKSP---DESLAEPKAPKA--LENSP 850
DB 549 --PAPAPPTPKLTS--ANPCPPVPAPANRRPAPAPAPAPAPAPAPAPAPAPAPAP 604
QY 851 KEQGVTTTTPAATKEDMTTAAKTKTERDLRTTPETTTAAPKMTETATTEKTESKI 910
DB 605 --PAPAPAPAPALPFPVNPAP-----PTTPAPK-----SRPAL 637
QY 911 TATTTQVSTTQDTTPEKITTLKTTTLAKVYTTTKITTEIMNKPEETAKPKDRAT 970
DB 638 PAAPAPAPAPVAVATTP-----PAPAPAPAP 665
QY 971 SKATTPKQKP---TKAPKPTSTKPKPTMPVRKPTTPPKMTSTM-----DELNP 1021
DB 666 SMALPAPAPDPPIPLATPAPAPAPPLPMSPPAPAPLPPAPAPAPAPPLPINOSSPLAP 725
QY 1022 TSLIAAMLTQTTTRPQOTNSKIVENPNPSEDAAGGEGET 1061

```

Db 726 VPGAPLAPLINGRPVPAKNSLI-----GSSSGDT 756

Search completed: April 26, 2002, 16:30:16
Job time: 631 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:19:45 ; Search time 210.7 Seconds

(Without alignments)
853.197 Million cell updates/sec

Title: AAI

Perfect score: 6568

Sequence: 1 MAMKTLPIYLLLSLVFVIQ.....AAITTRSGQTLKWMYNC 1229

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6470.5	98.5	1404	4	Q92954
2	6455.5	98.3	1404	4	Q92954
3	3173.5	48.3	1054	11	Q92954
4	1713.5	26.1	401	6	Q92954
5	939.5	14.8	1079	5	Q92954
6	939.5	14.3	1049	5	Q92954
7	865.5	13.2	1795	5	Q92954
8	853	13.0	1489	10	Q92954
9	808.5	12.3	1315	10	Q92954
10	802	12.2	1274	5	Q92954
11	800.5	12.2	1480	10	Q92954
12	784	11.9	1188	10	Q92954
13	778.5	11.9	2187	11	Q92954
14	739	11.3	2112	5	Q92954
15	735	10.9	763	2	Q92954
16	707.5	10.8	555	10	Q92954
17	671	10.2	4880	11	Q92954
18	667.5	10.2	4880	11	Q92954
19	667.5	10.2	5085	11	Q92954

20	663.5	10.1	2284	5	Q9VPG1	Q9VPG1 drosophila
21	660.5	10.1	6677	5	Q9N435	Q9N435 caenorhabd
22	659.5	10.0	1229	5	Q94185	Q94185 caenorhabd
23	653.5	9.9	1514	5	Q9GUV7	Q9GUV7 leishmania
24	647	9.9	4833	11	Q9QYX6	Q9QYX6 mus musculu
25	647	9.9	5038	11	Q9QYX7	Q9QYX7 mus musculu
26	643	9.8	2089	4	Q14676	Q14676 homo sapien
27	635	9.7	7962	4	Q10465	Q10465 homo sapien
28	633	9.6	489	10	Q41707	Q41707 vigna ungu
29	632	9.6	761	10	Q92410	Q92410 arabidopsis
30	632	9.6	6632	5	Q17362	Q17362 caenorhabd
31	629	9.6	3507	5	Q23587	Q23587 caenorhabd
32	628.5	9.6	1612	5	Q9VX02	Q9VX02 drosophila
33	627	9.5	990	13	Q91803	Q91803 xenopus lae
34	625	9.5	6642	5	Q01761	Q01761 caenorhabd
35	624.5	9.5	2768	5	Q9VC00	Q9VC00 drosophila
36	622.5	9.5	839	2	Q9RX57	Q9RX57 delinococcus
37	622	9.5	3570	4	Q9XVS4	Q9XVS4 caenorhabd
38	620.5	9.4	971	5	Q9XVS4	Q9XVS4 caenorhabd
39	615	9.4	2344	5	Q9N3Y8	Q9N3Y8 caenorhabd
40	607.5	9.2	801	5	Q23635	Q23635 caenorhabd
41	607.5	9.2	924	12	Q99307	Q99307 epstein-bar
42	605	9.2	379	5	Q27929	Q27929 drosophila
43	600.5	9.1	1893	5	Q9NKC9	Q9NKC9 drosophila
44	598.5	9.1	409	10	Q9SBM1	Q9SBM1 volvox cart
45	577	8.8	956	10	Q9LJ64	Q9LJ64 arabidopsis

ALIGNMENTS

RESULT	1				
ID	Q92954	PRELIMINARY;	PRT;	1404	AA.
AC	Q92954				
DT	01-FEB-1997 (TREMBLrel. 02, Created)				
DT	01-FEB-1997 (TREMBLrel. 02, last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)				
DE	MEGAKARYOCYTE STIMULATING FACTOR.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RN	SEQUENCE FROM N.A.				
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,				
RA	Fitzgerald M., Giannotti J., Calveti J., Fitzgerald M., Kriz M.J.,				
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,				
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;				
RL	Blood 78:279-279(1991).				
RN	[2]				
RN	SEQUENCE FROM N.A.				
RA	Meberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,				
RA	Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,				
RA	Jacobs K., Turner K.;				
RL	(In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,				
RL	Mosher D.F. (eds.);				
RL	Biology of vitronectins and their receptors., pp.45-52,				
RL	Elsevier Science Publishers B.V. (1993).				
RN	[3]				
RN	SEQUENCE FROM N.A.				
RP	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,				
RA	Kelleher K., Giannotti J., Calveti J., Fitzgerald M., Kriz M.J.,				
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,				
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;				
RL	Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.				
DR	EMBL; U70136; AAB09089.1; -				
DR	InterPro; IPR000585; Hemopexin.				
DR	InterPro; IPR001212; Synaptotagmin_B.				
DR	InterPro; IPR002400; GF_cysnct.				
DR	Pfam; PF00045; hemopexin; 2.				
DR	Pfam; PF01033; Somatostatin_B; 2.				
DR	PRINTS; PR00438; GFCYSKNCT.				

```

DR PRINTS: P000022; SOMATOMEDINB.
DR PROSITE; PS00024; HEMOPEPTIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

```

Query Match	98.5%;	Score 6470.5;	DB 4;	Length 1404;
Best Local Similarity	87.5%;	Pred. No. 0;		
Matches 1229; Conservative	0;	Mismatches 0;	Indels 175;	Gaps 11;

0Y	1	MAKTLPEYLLLLLSVFIQOVSS-----	24
Db	1	MAKTLPEYLLLLLSVFIQOVSSQDLSSCAGRGGEYSKDACNCNDYNQNHMECCPPE	60
0Y	25	-----	24
Db	61	KRVCAELSCGRGFESHEREDCDQAOCKKYKCCPDYESCAEYHNTPSSSKAP	120
0Y	25	-----	24
Db	121	PPSGASQTIKSTYKRSKPPKKTKKYIESELTTEHSVSEMOESSSSSSSSSTIM	180
0Y	25	-----YDNKNKNTKKKPPKPPVYVDAGSGLDNDGFKYTPDST	65
Db	181	KIKSSKNSAANRELQKTLAKVADKNKKNTKKKPPKPPVYVDAGSGLDNDGFKYTPDST	240
0Y	66	TOHNKVSISPKITTAKPINPNSPLPMSDSDSKESLTVNKEETVEYKETTITTKKONSTOG	125
Db	241	TOHNKVSISPKITTAKPINPNSPLPMSDSDSKESLTVNKEETVEYKETTITTKKONSTOG	300
0Y	126	KEKTSASAKETOSIEKTSAKDLAPISKVILAPTPKAETTTGKPALTPKEPTPTTKEPAS	185
Db	301	KEKTSASAKETOSIEKTSAKDLAPISKVILAKTPKAETTTGKPALTPKEPTPTTKEPAS	360
0Y	186	TTPKPEPTPTTIKSAPTPKKEBAPTTTYSAPTTKKEBAPTTTKKEBAPTTTKKEBAPT	245
Db	361	TTPKPEPTPTTIKSAPTPKKEBAPTTTYSAPTTKKEBAPTTTKKEBAPTTTKKEBAPT	420
0Y	246	APTTTKSAPTPKKEBAPTPPKKPAPTTPKKEBAPTPPKPEPTTPKKEBAPTTTKKEBAPT	305
Db	421	APTTTKSAPTPKKEBAPTPPKKPAPTTPKKEBAPTPPKPEPTPTTKKEBAPTTTKKEBAPT	480
0Y	306	EPAPTAAPKPAPTTPKKEBAPTPPKKEBAPTTTKKEBAPTTTKKEBAPTTTKSAPTTTKKEBAPT	365
Db	481	EPAPTAAPKPAPTTPKKEBAPTPPKKEBAPTTTKKEBAPTTTKKEBAPTTTKSAPTTTKKEBAPT	540
0Y	366	TTKSAPTPPKPEBPTTKKEBAPTPPKKEBAPTPPKKPAPTTPKKEBAPTPPKKEBAPTTTKKP	425
Db	541	TTKSAPTPPKPEBPTTKKEBAPTPPKKEBAPTPPKKPAPTTPKKEBAPTPPKKEBAPTTTKKP	600
0Y	426	APTAAREAPPTPKETAAPTPPKLTPPTPEKLAETTPPEKRAPTTPEELAPTTPEEPTPTT	485
Db	601	APTAAREAPPTPKETAAPTPPKLTPPTPEKLAETTPPEKRAPTTPEELAPTTPEEPTPTT	660
0Y	486	PEEPAPTPPKAAAPNTKKEBAPTPPKKEBAPTPPKKEBAPTPPKETAAPTPPKGATPPTIKKEP	545
Db	661	PEEPAPTPPKAAAPNTKKEBAPTPPKKEBAPTPPKKEBAPTPPKETAAPTPPKGATPPTIKKEP	720
0Y	546	APTTAKKPAKPAKLAETTTKEPTSTSDKAPATTPKGAATTPPKKEBAPTTKKEBAPTPPKG	605
Db	721	APTTAKKPAKPAKLAETTTKEPTSTSDKAPATTPKGAATTPPKKEBAPTTKKEBAPTPPKG	780
0Y	606	TAPPTLKKEBAPTPPKKPAKELAPTTTGGPSTTSDKAPATTPKEBAPTPPKKEBAPTPPK	665
Db	781	TAPPTLKKEBAPTPPKKPAKELAPTTTGGPSTTSDKAPATTPKEBAPTPPKKEBAPTPPK	840
0Y	666	KAPAPTPPEPPTTSEVSTPTTKKPTTIHKSPDESPTSELSABETPALENSPEEPVPT	725
Db	841	KAPAPTPPEPPTTSEVSTPTTKKPTTIHKSPDESPTSELSABETPALENSPEEPVPT	900
0Y	726	TTPAPATPEMTTAKDCTTEBDLTTPPETTTAAPKAMTKEATTTTEKTESKLTALTTOV	785

Db	901	TKTLPAAKPEKTTTAKDKTTERBRLTRTTPETTTAPAKMTKEFATTTTEKTTESKITTATTTQV	960
QY	786	TSTTODUTPEKITTLEKTTTLAERVTTTKYITTEIIMKKPEETAKPKDRATNSKATTPK	845
Db	961	TSTTODUTPEPKITTTLKTTTLAKRVTTTKYITTEIIMKKPEETAKPKDRATNSKATTPK	1020
QY	846	PQKTTAKPKKPTSTKRRKKTMPRVKRRKKTPTTPRKKMTSTMBELNPTSRIAEMALQTTTPBN	905
Db	1021	PQKPTKAPKKTSTKRRKKTMPRVKRRKKTPTTPRKKMTSTMBELNPTSRIAEMALQTTTPBN	1080
QY	906	QTPNSKILEVNPCKEDGAGEGETPHMLLRPHVMEVTPDDMDYLPVPNOGIIINPMLS	965
Db	1081	QTPNSKILEVNPCKEDGAGEGETPHMLLRPHVMEVTPDDMDYLPVPNOGIIINPMLS	1140
QY	966	DETNICNGKPYDGLTTLRNGTTLVAERNGHYFWMLSPSPSPSPARRITEWGCIPSDIYFT	1025
Db	1141	DETNICNGKPYDGLTTLRNGTTLVAERNGHYFWMLSPSPSPSPARRITEWGCIPSDIYFT	1200
QY	1026	RCNEGKTEFFPKDSQYKRFNNDIKDAGYPRPIFKGGGLTGOIVALSTAKKMMPE5Y	1085
Db	1201	RCNEGKTEFFPKDSQYKRFNNDIKDAGYPRPIFKGGGLTGOIVALSTAKKMMPE5Y	1266
QY	1086	FFPKGGSTGOQYIKQOEVOVKCPGRPALNTPVYGEMTQVRRRREFERALISQSTHTRIQY	1145
Db	1261	FFPKGGSTGOQYIKQOEVOVKCPGRPALNTPVYGEMTQVRRRREFERALISQSTHTRIRIQY	1320
QY	1146	SPARLAYODKGVLLHNEKVSILMKGLDPNVVTSALSPNTRKPDGXYDYAFASKDQYYNIDY	1205
Db	1321	SPARLAYODKGVLLHNEKVSILMKGLDPNVVTSALSPNTRKPDGXYDYAFASKDQYYNIDY	1380
QY	1206	PSRTARATTTSSGOTLSKVMYNCP	1229
Db	1381	PSRTARATTTSSGOTLSKVMYNCP	1404

RESULT	2	
Q9BX49		
ID	Q9BX49	PRELIMINARY; PRT; 1404 AA.
AC	Q9BX49;	
DT	01-JUN-2001 (TREMBLrel. 17, Created)	
DR	01-JUN-2001 (TREMBLrel. 17, Last sequence update)	
DI	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE	BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR).	
GN	BG174L6.2.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Wray P.	
RL	Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.	
SR	EMBL; AL13553; CAC36090.1; -	
SQ	SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEE5 CRC64;	

Query Match	Similarity	Score	DB	Length
Best Local	87.3%	Pred. No. 0;		
Matches 1226;	Conservative	0;	Mismatches	3; Indels 175; Gaps 1.
QY	1 MAMKTLPIYLLLLLSVFVITQOVSS	-----	24	
Db	1 MAMKTLPIYLLLLLSVFVITQOVSSQDLSGACGCGSGTSDATCNDYMCQHYMECCPDF	-----	60	
QY	25 -----		24	
Db	61 KRYCTAELSCGRCFESFERGREGDCDAQCKRYDKCCPDYESFCAEVHNPTSPSSKAP	-----	120	
QY	25 -----		24	
Db	121 PEGASQITKSTTKRSKPRPNKKTKTKYIESEBITTEHSVSEMOESSSSSSSSSSSTIR	-----	180	

OY	25	-----XKDKKNNRTKKKKPPKRPVYVDEAGSGJDNDEKVTVPDST	65
Db	191	KIKSGKNSAANRELQKLLKAKDKNNKTKKKPPKPPVYVDEAGSGJDNDFKVTVPDST	240
OY	66	TQHNKVSPTITAKPAPINRPSLPSPDSKSTSLTVNKEKTVVEIKETTTNNKOTSIDG	125
Db	241	TQHNKVSPTITAKPAPINRPSLPSPDSKSTSLTVNKEKTVVEIKETTTNNKOTSIDG	300
OY	126	KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEBPTTPKEBAS	185
Db	301	KEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEBPTTPKEBAS	360
OY	186	TTPKBEPPTTTKSAPTPKBPAPTTTKSAPTTPKEBAPTTPKEBAPTTPKEBAPT	245
Db	361	TTPKBEPPTTTKSAPTPKBPAPTTTKSAPTTPKEBAPTTPKEBAPTTPKEBAPT	420
OY	246	APTTPKSAPTTPKBPAPTTTKBPAPTTTPKBPAPTTKEBPTPTTPKBPAPTTKEBAPT	305
Db	421	APTTPKSAPTTPKBPAPTTTKBPAPTTTPKBPAPTTKEBPTPTTPKBPAPTTKEBAPT	480
OY	306	EPAPTAAPKBPAPTTPKBPAPTTTPKEBAPTTPKEBPTTPKEBAPTTPKSAPTTPKEBAPT	365
Db	481	EPAPTAAPKBPAPTTPKBPAPTTTPKEBAPTTPKEBPTTPKEBAPTTPKSAPTTPKEBAPT	540
OY	366	TTKSAPTTPKBPBPTTTKEBAPTTPKEBAPTTPKKAAPTTPKEBAPTTPKEBAPTTPKBP	425
Db	541	TTKSAPTTPKBPBPTTTKEBAPTTPKEBAPTTPKKAAPTTPKEBAPTTPKEBAPTTPKBP	600
OY	426	APTAPKEBAPTTPKEBAPTTPPKLPTTPPKLAPTTPKEBAPTTPBELAPTTPBEPPTPT	485
Db	601	APTAPKEBAPTTPKEBAPTTPPKLPTTPPKLAPTTPKEBAPTTPBELAPTTPBEPPTPT	660
OY	486	PEBPAPTTPKKAANNTPKBEPAPTTTPKEBAPTTPKEBAPTTPKETAAPTTKGPAPTTLKEP	545
Db	661	PEBPAPTTPKKAANNTPKBEPAPTTTPKEBAPTTPKEBAPTTPKETAAPTTKGPAPTTLKEP	720
OY	546	APTTPPKKAPKAPKELAPTTTKBPTSTSDKAPATTPKGTAPTTPKBPAPTTPKBPAPTTBKG	605
Db	721	APTTPPKKAPKAPKELAPTTTKBPTSTSDKAPATTPKGTAPTTPKBPAPTTPKBPAPTTBKG	780
OY	606	TAPTTLKEBAPTTPKKAAPKAPKELAPTTTKGPTSTSDKAPATTPKETAAPTTTPKEBAPTTPK	665
Db	781	TAPTTLKEBAPTTPKKAAPKAPKELAPTTTKGPTSTSDKAPATTPKETAAPTTTPKEBAPTTPK	840
OY	666	KPAPTTPETPTPTTSEVSTPTTTKEPTTTIKSPDESSTPELSAEPPTKALENSKAPBAPT	725
Db	841	KPAPTTPETPTPTTSEVSTPTTTKEPTTTIKSPDESSTPELSAEPPTKALENSKAPBAPT	900
OY	726	TKTPAATKPEWTTAKDITTERDLRTTPETTTAAPKMTKEIATTTTEKTTESKITATTQV	785
Db	901	TKTPAATKPEWTTAKDITTERDLRTTPETTTAAPKMTKEIATTTTEKTTESKITATTQV	960
OY	786	TSTTTQDTPPKITTLTKTTLAKRVTTTKTTTTEIIMKPREIAPKPDRAIINSKATTPK	845
Db	961	TSTTTQDTPPKITTLTKTTLAKRVTTTKTTTTEIIMKPREIAPKPDRAIINSKATTPK	1020
OY	846	POKPTKAPKKTSTKPKKPTPVPRVKPTTPTPPKMSTMBELNPTSRILAEAMLQTTTPBN	905
Db	1021	POKPTKAPKKTSTKPKKPTPVPRVKPTTPTPPKMSTMBELNPTSRILAEAMLQTTTPBN	1080
OY	906	QTPNSKLEVNPKSEDAGAGEGTPHMLLRPHVMEVTPPDMDYLRPVNOGIINPMLS	965
Db	1081	QTPNSKLEVNPKSEDAGAGEGTPHMLLRPHVMEVTPPDMDYLRPVNOGIINPMLS	1140
OY	966	DETINICNGKPYDGLTILRNGTILVAFGHYFWMLSPPSSPSPARRITEWGISPIDIYFT	1025
Db	1141	DETINICNGKPYDGLTILRNGTILVAFGHYFWMLSPPSSPSPARRITEWGISPIDIYFT	1200
OY	1026	RCNCEGKTEFFFKDSQVYRPFINDIKDAGYRPIKJGSGGLTGOIYVALSTAKYKMMPEVY	1085
Db	1201	RCNCEGKTEFFFKDSQVYRPFINDIKDAGYRPIKJGSGGLTGOIYVALSTAKYKMMPEVY	1260
OY	1086	FFKRGSGIIOOYIKQEPVOKPCGRBAPLNPVYGEVTOVRRRRFERAIGPSQTHIRIOY	1145

Db	1261	FFKRGSGTQQYTYKCEPVQKCPGRPALNTPYGGTTQVRRRRREREAIGSPQTTIRIÖY	132
Qy	1146	SPARLAYODKGLHNEHVVYSILMRCLPWVWVSATSLPPIRKROGDDYAFSKDDQYINIDV	120
Db	13321	SPARLAYODKGLHNEHVVYSILMRCLPWVWVSATSLPPIRKROGDDYAFSKDDQYINIDV	138
Qy	1206	PSRTARAATTTRSGOTLSKRWYNCP	1229
Db	1381	PSRTARAATTTRSGOTLSKRWYNCP	1404
RESULT	3		
Q9JMG9		PRELIMINARY; PRT; 1054 AA.	
ID	Q9JMG9		
AC	Q9JMG9;		
DT	01-OCT-2000 (Tremblrel. 15, Created)		
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)		
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)		
DE	MRNA, COMPLETE CDS, SIMILAR TO MEGACARBOCYTE STIMULATING FACTOR PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN. PRG4.		
GN	PRG4.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N. A.		
RA	Ikegawa S., Nakamura Y.;		
RT	"a novel mouse gene highly homologous to a human gene encoding megakaryocyte stimulating factor precursor and cartilage superficial zone protein.";		
RT	Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.		
RL	EMBL: AB044730; BAA92310.1; -		
DR	MGP; MGI:1891344; Prg4.		
DR	InterPro; IPR000585; Hemopexin.		
DR	InterPro; IPR002965; P_rich_extensn.		
DR	InterPro; IPR001212; Somatomedin_B.		
DR	Pfam; PF01033; Somatomedin_B. 2.		
DR	PRINTS; PR01217; PRICHEXTENS.		
DR	PRINTS; PR00022; SOMATOMEDINB.		
DR	SMART; SM00120; HX; 2.		
DR	SMART; SM00201; SO; 2.		
DR	PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.		
DR	PROSITE; PS00524; SOMATOMEDIN_B; 2.		
DR	SEQUENCE 1054 AA; 115991 MW; 4FC64BPA4283235 CRC64;		
Qy	19	IQGVSSVKDNKKNRTKKRPKPYVDVDEAGSGLDNGDEKVT--TPDSTTQHNKVS	76
Db	189	LQKNNVNDKNKKRPKPYVDVDEAGSGLDNGDEKFLPPDPDPTPHSKVATSPK	248
Qy	77	ITTAELINPROSLPNSTKRETSITVYKKETTVETKETTNNKQTSIDGKKKTSAAEQ	133
Db	249	TTAAAPVVPKPSLANSETSKEASIASAKETTVETKETTANNKQSSA-SKKKTSVSETR	307
Qy	137	SIEKTSADLDAPTSKVLAKPPTKAEITTKGPAITTPKEPPTTPKPEASTPKKEPPTTI	198
Db	308	SAEKTSQKDV-----EPSTTPK-----	322
Qy	197	KSAPPTKPEAPPTTKSAPTTPKPEAPPTTKPEAPPTTPKPEAPPTTKSAPTT	256
Db	326	NSAPPTTKKPV-TTKESKFLP-----LQGEDEPTTAKEPPTTKKPEPTT	370
Qy	257	PKPEAPPTTKKPEAPPTTKPEAPPTTKPEPTPTTPKPEAPPTTKPEAPPTTKPEAPPA	316
Db	371	RKEPPTTKPEDEPPTTKPEPPTTKPEPPTTKPEPPTTKKRPPTTKPEGPTTKPE	430
Qy	317	PTTPKEAPPTTKPEAPPTTKPEPPTTKPEAPPTTKSAPTTKPEAPPTTKSAPTTKPE	376

RT "Genome sequence of the nematode *C. elegans*: a platform for
RT investigating biology. The *C. elegans* Sequencing Consortium.",
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Colton M.;
RT "The sequence of *C. elegans* cosmid Y51B11A."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006797; AAF60743.1; -
DR InterPro: IPR002965; P-rich extensin.
DR PRINTS: PR01217; PRICEXTENSX.
SQ SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;

Query Match 14.8%; Score 969; DB 5; Length 1079;
Best Local Similarity 29.9%; Pred. No. 6.3e-57;
Matches 327; Conservative 137; Mismatches 472; Indels 156; Gaps 44;
QY 22 VSSVKNKKNNKKKKPKPPVVDGAGSLDNDGFKVTT--PDSTTQHKKVSTSP-KI 77
DB 62 LSSPTSSSTPIKETTATTAETSTPESSSTTPVQTTTAAETSTTEAPSSSTTPVQT 121
QY 78 TLAKPIPRSLPNSDTS--KETSIVNKETIVERKETTNNKOSTDCKEYTSKAKET 135
DB 122 TTTTAAETSTPESSSTSPVQTTTAAETSTTEAPSSSTTPVQTTTAAETSTTEAP 175
QY 136 QSIETKSADKLAPTSKVLAKPTPKAETTTKGP--ALTTPEPTPTPKKEPASTPK 193
DB 176 TSTEPSSS---TSVQTTTAAETSTPESSSTTPVQTTTAAETSTTEAPSSST 231
QY 194 TTTSKAPTTPKEPAPTTTTSKAPTTPKEPAPTTTKEPAPTTTKEPAPTTT 253
DB 232 TPVQTTTAAETSTPESSSTTPVQTTTAAETSTPESSSTTPVQTTTAAETST 287
QY 254 PTTPKEPAPTTPPKPAPTTPKEPAPTTPPKPAPTTPKEPAPTTPKEPAPTTP 312
DB 288 PETSTPESSSTTPVQTTTAAETSTTEAPSSSTTPVQTTTAAETSTTEAPSS 345
QY 313 KKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSA 370
DB 346 -----TTTPVQTTTAAETSTPESSSTTPVQNTTTAAETSTTEAPSSSTTP 396
QY 371 PTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 430
DB 397 TTTTAAETSTPESSST--TPVQTTTAAETSTTEAPSSSTTPVQTTTAAETST 453
QY 431 KEP--APTTPKEPAPTTPPKLPTTPPEKLAPTTPKEPAPTTPPELAPTTPPE 486
DB 454 TEPPSSSTTPVQTTTAAETSTTEAPSSSTTPVQTTTAAETSTTEAPSSSTTP 509
QY 487 EEPAPTTPKAAPNTPKKEPAP--TTTPKEPAPTTPKEPAPTTPKEPAPTTPK 543
DB 510 VQTTTAAETSTPESSSTTPVQTTTAAETSTTEAPSSSTTPVQTTTAAETST 564
QY 544 EPAPTTPKAPAPKELAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 600
DB 565 ETSTPESSSTTPVQTTTAAETSTTEAPSSSTTPVQTTTAAETSTTEAPSS 621
QY 601 TTPKGAPTTLKEPAPTTPPKKAPKELAPTTPKGTSTSDKP--APTTPKEPAPTTP 659
DB 622 TTPVQTTTAAETSTPESSSTTPVQTTTAAETSTTEAPSSSTTPVQTTTAAET 681
QY 660 APTTPKAPAPTTPPEPTTPSEVSTPTTKEPTTHKSDESTPELSAPPTPKALNS 719
DB 682 TSTEP---PSSSNTPVQTTTAAETSTTEAPSSSTTPVQTTTAAETST----- 731
QY 720 EBGVPTT--KTDAATKPEWTTTAAKDKTTPBDRLRTT---PETTTAAAPKATK 773

DB 732 PPSSSTTPVQTTTAAETSTTEAPSSSTTPVQTTTAAETSTTEAPSSSTTPVQTTT 791
QY 774 TESKITTATTTQVNTSTVQDTPPKITTLTKTTTAAVTTTAKTTTTEIMNKKEETA-KP 832
DB 792 TAPETSTPESSSTTPVQTTTAAETSTTEAPSSSTTPVQTTTAAETSTTEAPSS 847
QY 833 KDAATNSKATTPPKPQKPTAPKPKPTSTKPKTMPRVKPKPTTPPKMTSTPELNPTSR 892
DB 848 PSSSTSPVQTTTAAETSTTEAPSSSTTPVQTTTAAETSTTEAPSSSTTPVQTTT 907
QY 893 IAEAMIGTTTRAPQTPNSKIVENPKSEDAAGNEG--ETPMMLLRHVAFVETP---DMD 948
DB 908 APET---TSTEP--PSSSTTPVQTTTAAETSTTEAPSSSTTPVQTTTAAETST 961
QY 949 YLPRV-----PNOGIINPMLSDETNT-----ICNGKPVGLTTP 981
DB 962 YIDRVVYPTTEWEMEKROIIGSYDSPRRTAESFVSTEDIGCTAALICITYSEGGIN 1021
QY 982 LRNGTLVARGHAFWMLS-----PSPSPPARITLEVWGIDPSIDVTRCNEGKTF 1034
DB 1022 L-NATL-----FGLSDGSSSIDLPF-----YVNPGLGE-IMPEINCEGKN- 1059
QY 1035 FFKDSQYWRFTN 1046
DB 1060 -----MSYNN 1064
RESULT 6
Q917S1 PRELIMINARY; PRT: 1049 AA.
AC Q917S1
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-JUN-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG53228 PROTEIN.
GN CG53228.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chertky J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dudin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegwam C.,
RA Jatala M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacלב J.M.,
RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 RN (12)
 RP SEQUENCE FROM N.A.
 RA Cadien E., Dreano S., Lelaure V., Mottier S., Gallibert F.;
 RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
 RL Submitted (Jul-1998) to the EMBL/Genbank/DBJ databases.
 RN [13]
 RP SEQUENCE FROM N.A.
 RA Benos P.;
 RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF003421; AAF4564.1; -;
 DR EMBL: AL031028; CA19845.2; -;
 DR Flybase: FBgn0025390; EG:5667.1;
 DR InterPro: IPR002557; Chitin_binding.
 DR InterPro: IPR002965; P-rich_extensin.
 DR Pfam: PF01607; Chitin_bind_2; 2.
 DR PRINTS: PRO1217; PRICHEXTENSIN.
 DR SMART: SM00494; ChEBD2; 2.
 SQ SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match 13.2%; Score 865.5; DB 5; Length 1795;
 Best Local Similarity 28.5%; Pred. No. 9.3e-50;
 Matches 276; Conservative 129; Mismatches 447; Indels 115; Gaps 28;
 QY 30 KNRKKKPPKPPVVDAGSGLDNGDFKVTTPDTSTTQHNKSTSPKITANKPINRPSL 89
 DB 431 KNDVEAPRIKSP-----KGLHLSNIVILPETTTT---TTTTKRVILCTPISPPDT 483
 QY 90 PMSD-----SKETSLVNEKETTVEKETTNNKOTSDGKEKTSAKETOSI 138
 DB 484 PPSSTAVATKSTPKISSTEOHSTTAKTTRKPRVTEKISSATEKRRTVAVTTTQK 543
 QY 139 ERTSADLAPTKSVLAKPRKAEITTKGPAITPKKEPTTPPKERASTKREPTTIKS 198
 DB 544 RSTTHNHSPPDKTIRSTLTSPKTTTPTSTTPTST-----TPTSTTPTST--- 594
 QY 199 ADTPKEAPATTKSAPTTKKEAPATTK-EPAPTTKKEAPATTKKEAPATTKSAPTTP 257
 DB 595 --TPTSTPTSTTIVKSTHPRRTTSQKTTASTTKK---TTPSPKTTTITDIPIST 648
 QY 258 KRAPATTKKRAPATTKKEAPATTKKEPTTTPKKEAPATTKKEAPATTKKRAPATTKK 317
 DB 649 TSKLSTTQKTTTTHKETAATSTSEKPRTEKTSVSTTKKSTESSPKSTSGKPT 708
 QY 318 TTPKKEAPATTKKEAPATTKKEPSPTTKKEAPATTKSAPTTKKEAPATTKKAPATTKPEP 377
 DB 709 TTPKSTRTPTTTPKVTITTOITTTTPLRSSSTETSTQEPPTTTTQPTTTTTLIVTPKTS 768
 QY 378 SPTTKKEAPATTKKEAPATTK-----KRAPTTKKEAP-----TTPKE 416
 DB 769 TTTTTEKRITSSPKPTTTOKTSTAPRTKVALITQKETTPQSTSTTITRKTNN 828
 QY 417 PAPTTKEAPAPKKEAPATTKP-ETAPTTPKKLPTTPKEKLAAPTTPKPEP-----AP 467
 DB 829 PEPSTTEKPTISTPKPSTTPKSTIVASSTETKISSPKPTTEKSTEPSTPNVSKSL 888
 QY 468 TTPKEAPATTKKEAPATTKPEEPATTKPEAPATTKAAAPNTPKKEAPATTKKAPATTK-TPKEAPAT 525
 DB 889 TSSQRASTSTSEPTKT-ONTTTPPKPTLKSTQELTSTOKVSVITTTTKATES 947
 QY 526 PKETAPTPKGAPTTKLEAPATTKKAPKELAPTTTKEPT--STSDKAPATTPKGA 583
 DB 948 PLTISTEENPTPKPLKPTTPTTSVATRTITTTTISESTETSTQKPKSTPTSTT 1007
 QY 584 PTPKEAPATTKKEAPATTKGATPTTKLEAPATTKKAPKAPKELAP--TTTGCPSTSDK 642
 DB 1008 RTTPKAVTVIVSTQNPPTTTTSTSTVTI-----TTP-NPSSTQRPRTTTPKOPTISITAST 1061
 QY 643 PAPTPPKETAPTPKKEAPATTKKAPATTKPETTPPTSEVSTPTT-TKE----- 690

DB 1062 TSGITRIPPTTNNPQNSTSTDLTIVTRPCPDPOSTSDKNNTACTOGLQVNIIELOS 1121
 QY 691 -----PTTHKSPDESTPELSAEPYKALENSKEFGVPTTKPAATKPEMTTAKD 742
 DB 1122 POKOEFTHTRHTALTGSRNLTGGOEVPDYMDAPSSAEASGCAWTAKAPMTSLAA 1181
 QY 743 KTTT-----DLRTPPETTTAPKMKETATTEKTESKITATTTQVSTTTQDTPPKIT 799
 DB 1182 HLLQKLFHISTTPREHAP--TORPSSQSSQSRK-GVITAOVARINLATSKPFLAH 1238
 QY 800 TLKTT---TLAPKYTTTKITTTTEIMNKPEETAKPKDRAATNSKATPPKPKPA-PK 854
 DB 1239 SLRSTIOQLASTQKRSIPKKTIVHTNTKPEPD--SEYDSETSEQYTDNDNVLKTOPR 1297
 QY 855 KPTSTKPKPTMRYKPKTKPTTPPKKTSMPLELNPTSRKLAEMLOTTRPNQTPNSKLYE 914
 DB 1298 AMSSTVAALVAVBSTTTEREPK-TSSSP---SPN---KATSTTQPIETTTGDLXY 1350
 QY 915 VNPKSED 921
 DB 1351 DSSGSSD 1357

RESULT 8
 096449 PRELIMINARY; PRT; 1489 AA.
 ID 096449;
 AC 096449;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
 GN CARGO.
 OS Phytophthora infestans (Potato late blight fungus).
 OC Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
 OC Phytophthora.
 OX NCBI_TaxID=4787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RACE 1-11;
 RA Goenhardt B.;
 RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF061185; AAC72308.1; -;
 SQ SEQUENCE 1489 AA; 164037 MW; 764AC79D1G2F5163 CRC64;
 Query Match 13.0%; Score 853; DB 10; Length 1489;
 Best Local Similarity 32.3%; Pred. No. 5.3e-49;
 Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;
 QY 33 TKKKPPKPPVVDAGSGLDNGDFKVTTPDTSTT-----OHNKVSTSPKITTAKPINR 86
 DB 338 TPYAPTEKPYDVEETTYTBEETVAPTKSETNAPTERMHYAHIEPCDEVTVMVAPTEET 397
 QY 87 PSLP-----PMSD-----SKETSLVNEKETTVEKETT---TNKQSTSD 124
 DB 398 TYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYAPTEETTYAPTEETTYAP 457
 QY 125 GKEKTSAKENQSIKTSADLAPTSKVLAKPTKAEETTTGPAITTKKEPTTPPKPEPA 184
 DB 458 TEETTYAPTEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 514
 QY 185 STTPKEPTPTTKSAPTTKKAPATTKKAPATTKKAPATTKKAPATTKKAPATTKKAPATTK 241
 DB 515 EETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 573
 QY 242 TKKEAPATTKKAPT-----TPKKEAPATTKKAPATTKKAPATTKKAPATTKKAPATTK 291
 DB 574 TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 632
 QY 292 P-----AP---TTKEAPATTKKEP-----APTAAPKAPATTKKAPATTKP 328
 DB 633 PTEETTYASTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 692

QY 329 KEPAAT--TTKEPSPPTKEP-----APT--TTKSAPT-----TTKEP-----A 363
 DB 693 TTYAPTEETTYAPAETTPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 752
 QY 364 PT--TTKSAPT--TPKEPSPPTKEP-----APTTPKEPATTPPKKAPPTTPKEPAT-- 412
 DB 753 PTEATTYAPTEETTYAPAETTPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 812
 QY 413 TKREPATTTKKAPAPAPAPAPATTPKETAFTPKKLTTPPKLAPPTPEKAPAPTPPEE 472
 DB 813 TTYEPTTEETTYAPTEETTPYEET--TYTPTEETTYAPTEETTYAPTEETTYAPTEET 870
 QY 473 LAPT--TPPEPTTPPEEP-----APT-----TPKAPAPATTPKEPATTPKEP- 513
 DB 871 YAPTEETTPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYEPT 930
 QY 514 ----APTTPKEPATTPKETAFTTPPKGAPT--TKKEPATTPPKKAPAP--ELAPT--TT 563
 DB 931 EETTYAPTEETTYAPAETTPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 990
 QY 564 KEPTSTSDKAPATTPKGAFTTPKEPATTPKEP-----APTTPKGAFT----- 609
 DB 991 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1050
 QY 610 -----TKKEPATTPPKP-----APKE--LAPT--TTKGPTSTSDKAPATTP 648
 DB 1051 EETTYAPTEETTYAPAETTPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1110
 QY 649 KETAPTPKEPATTPPKKAPATTPETPPPTTSVSTPTTKEPT--TIKSPDESPELS 706
 DB 1111 YAPAEETTPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1169
 QY 707 AETTPKA--LENSPKEG-----VPTTKTPAATPKEMTTAKOKTREDLKTPT-----ETT 756
 DB 1170 TETTPAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYEPT 1229
 QY 757 TAAPKTKETATTTKTESKITATTTQVSTTTODTT-----PEKITTT--LKT 803
 DB 1230 TYAP-----TEETTPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1284
 QY 804 TTYLAPVTTTKITTTTEIMNKPEETAKPKDRAT--NSKATTPKP--OKTPKAKKPT--- 857
 DB 1285 TTYAPTEATTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1344
 QY 858 -----STKRP--KTMPIVRKPTTPPKKMTSTMPBELNPSRLAEAMLQTTTP--N 905
 DB 1345 PAESTSTVSTEKPCNTEETDEPTDEPTDE--PSDEPTDEPTDEPTDLPTDEPTPCON 1402
 QY 906 QTPNSKLVEVNPKESEDAG 923
 DB 1403 QGINGIGVENKVRNNAG 1420

RESULT 9
 Q9SPMO PRELIMINARY; PRT: 1315 AA.
 AC Q9SPMO; 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
 DE EXTENSIN-LIKE PROTEIN.
 GN PEX2.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogonaceae; Zea.
 NC NCB1_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=POLLEN;
 RA Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
 R Hochorst D., Gao M., Showalter A., Bedinger P.A.;
 RT "Pollen Extensin-like (Pex) Genes in a Monocot and a Dicot.";

RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF159297; AAD55980.1; -
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR001998; Xylose_isom.
 DR InterPro: IPR002965; P-rich_extensin.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 3.
 DR PRINTS: PR01217; PRICHEXTENS.
 DR PROSITE: PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
 DR SMART: SM00370; LRR; 3
 SQ SEQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;

Query Match 12.3%; Score 808.5; DB 10; Length 1315;
 Best Local Similarity 25.0%; Pred. No. 4,4e-46;
 Matches 255; Conservative 88; Mismatches 459; Indels 219; Gaps 33;

QY 37 PPP-----KPPVVDAGSGLDNGDFKYVTPDSTTHNNKYSTSKITTAKPINFR 86
 DB 412 PTPGPPSSSVKPPPSVPG-----KPAAPAPMPTPTPTPDVSP-----EPL-PE 455
 QY 87 PSLPP-----NSDSKETSLVKNKETTYEKEETTTTKQNSTDGEKETSARETO 136
 DB 456 PSEVPAPAPMPTLRSPPADEYIPT-----PPVAPSPGTSPPASRGAPPLQAOAPPAS 511
 QY 137 SLEKTSAKDLAPTSKY-----LAKPTKAEETTKGPAITTPKEPTTPPKE 182
 DB 512 SPAPATPVKSSPPAAVVLPPAKTSPAPVASSPPEAVSSPQGVKSPPPAPVASSP 571
 QY 183 PASTPKEPTPTTKAPATTPKEPATTPKSAFTTP--KEPA-----TTTKEPATTP 234
 DB 572 PPKKSP--PPAPVASSPPLMKSSPPAPVASSPQGLKSPPPVLMVSTPSVKSPPVPV 629
 QY 235 KEPAATTPKEPATTPKSAFTTPKE-----PAP--TTP-----KKPATTPKEPATTPKE 283
 DB 630 VASPPPVKSSPPPLAVSSPSPPVKPLPLPAPGKSTPPPEEKPTPTPVKSSPPPEK 689
 QY 284 PPTTPKEPATTPKEPATTPKKEPATTPKKEPATTPKKEPATTPKKEPATTP 343
 DB 690 PPTLTTPSPQEKPPSPSPSKPPSPSEVELTPPSKSSPPEEPVSSPPQAVKSSSP-- 747
 QY 344 KPEPATTTKSAPTTPKEPATTTKSAPTTPKPSPTTTTKEPATTPKKEPATTPKKA 403
 DB 748 ---PAPVS--SPPLKSSPPVPSSPPTPKSSPLAVSSPQVEKTSPPAPVSSP 802
 QY 404 TTPKEPATTPKEPATTTTKKAPAPAPAPATTPKETAFTTPPKKLTTPTEKLAFTPE 463
 DB 803 PTPKSSPLAPVSSPPQVEKTSPPAPVSSPPTPKSSPLAPVSSPQVEKTSPPAPV 862
 QY 464 KPAFTTPEELAPTPPEEPPTTPPEEPAPTPPKAAANTKEKAPATTPKKEPATTPKEPA 523
 DB 863 SSPPLPKSSPPSSSVSSPPTIVKSSPPAPPLSSPMTPKSSSPPAHVSSPPAEKSSSP 922
 QY 524 TTPKEPATTPKGTAT-----PTTKEPATTPPKKAPAKELAPTTTKEPTSTSD 572
 DB 923 LAPSSPSEPKSPSPPMWEKTSPPATVSSPPTPKSSPP--APVSSPPPVKSSP 979
 QY 573 KPAFT-----TPKGAFTTP-----KEPATTPKEPATTPK--GAFTTPKKAFT 617
 DB 980 PPAPVSSPPTPKPLPPAPVSSPPPVKSSPPTPVSSPPPTPKPLPPPVSSPPT 1039
 QY 618 TPKKPAKELAPTTTGPSTSDKRAPPT--KETAFTTPK--EPAFTTPKKAFT 671
 DB 1040 XKPLPP--APVSSPPPVKSSPPAPVSLPPTPKSPPKRVSSPPRVKCCPPPL 1096
 QY 672 PPTTPPTTSVSTPTTKEPTTTIKSPDSTPELSAFTPKALENSPKKEGVTTPPA 731
 DB 1097 VSSPPAPKPLPPPTPVSS-----PPVKSSTPTPVSSPPAPKSSPPTPV 1146
 QY 732 TKPEMTTAKDCTTERDLRTTP-----ETTPAAPKMTKETATTTETTSKITATTTQV 786
 DB 1147 SPP-----ELKSSPPAPVSSPPASPKSSPPAPVSLPPEVKSSPPAPIS 1194

```

QY 787 STTTODTTPKITTLLKTTTLAPKVTTKKTTTTEIMNKDEETAKPRDRATNSKATTPK 846
DB 1195 SPPPAKSP-----PPPA-----PMSLPPPKSPPPAVSSP 1228
QY 847 QKPTKAKKPTSTKKTMRVRKPKTTPPKMTSTMPPLNPTSLAEALQTTTPNQ 906
DB 1229 PPPKSSPPPAJSSPPPA--VKPSPPPAPVSSPPPAV-----TSAP-- 1271
QY 907 TPNSKLEVPKSEDAAGAGETPHMLRPHVMEVTPMDYLPRVNOGIITPMLS 966
DB 1272 -----PKKEEDSTA-----PRAELPPPSFDIILPIMAN 1302
QY 967 E 967
DB 1303 K 1303

RESULT 10
Q20007 PRELIMINARY; PRT; 1274 AA.
AC Q20007;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE COSMID F35A5.
GN F35A5.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabdilitida; Rhabdilitoidea;
OC Rhabdilitidae; Poloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RC MEDLINE=94150718; Pubmed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berts M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Crawford M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Showkeen R.,
RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thelery-Mleg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spirot J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN (12)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Leimbach D.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN (13)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN (14)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U46675; AAB52641.1; -.
SQ SEQUENCE 1274 AA; 138065 MW; 07F6BD292C1799F CRC64;

Query Match 12.2%; Score 802; DB 5; Length 1274;
Best Local Similarity 28.0%; Pred. No. 1.2e-45;
Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;
QY 30 KNRKAK-KP-----TP-----KPPVDEAGSGDNGD-----FKYTTDTSTGTQHNKYSTS 74
DB 274 KNPTKKWKPPWEDETPEVEVKEPEVPKAPVLLKKDPAPAAKARDSPSKAAKVPKPS 333

```

```

QY 75 PKITTAAP I-NPRSLPNSDTSKETSLTVNKETVETKETTNTKOTSTDGKEKTSK 133
DB 334 SPVPPPTPVKNPVKKYKPPMEVDEDEPAEVEKKSPAPKPKTPVLKRRKPESSSTTPSSDPS 393
QY 134 ETOSIEKTSKADLAPISVKVLAKPTPKAE---TTTKGPA-----LTTP 172
DB 394 PKRAAPAVKRRDSSSPKATLQADPKQAEVPPPVKNPVKKYKPPMEVDEDEPAEVEKOP 453
QY 173 KEPTTPTP-----KEPATTTKEP-----TP--TTKSAPTTKEPATTTKSAPTTPK 223
DB 454 EAPAKKTPVLKREPAKDTAKPATSKTPEPEKDDPVKRRDSSPKVAAKPOSAQAPA- 512
QY 224 TTKEPA-----PTTKEPATTTKEPAT---TTTKEPATTT 257
DB 513 TPVKNPVKKMRPPWEDETPADVSKTDAKKTTSKAKDPAKESLAKRADTKAAKAP 572
QY 258 KEP-----APTPKAPAPTPPK-----PAPTPKEPPTTPKEPATTTKEPATTPK-EP 307
DB 573 RDSPPKVAAPTADEKTPVLAKKEPAGPADSKTKEPEKSPRDSPPKAVPAKVPKTEV 632
QY 308 APTPAKKAP-----TTKEPATTTKEPATTTKEP-----SPTTKEPATTTKAP 357
DB 633 APAVAKKPEPISAPKDTAPKAEPNSPVP--PVPVKNPVKKMRPPWEDDAPAKVSLPE 691
QY 358 TTKEPATTTKSAPTTP--KEPSPPTTKEPATTPPK--EPAPTPKAP--TTTKEPAT 412
DB 692 PEKK--TVLAKKAPATKRDSEADPVSQSPSSKODKLAKKAPVKKRDSPPKAVPINKAPK 750
QY 413 TPKEPATTTKAPAPTA--PKEPATTPKETAPTPPKLPTTP-----EKLAPTPKPA 466
DB 751 T--EVPPAVVKKPEPVAKSDPSEKKA--AEPNSP--VVPPTPVKNPVKKMRPPWEDDA 805
QY 467 PTPEELAPTPTEEP-----TPTPEEPAPTPKAAAPNPKKEPATTPKEPATTT--PKE 520
DB 806 PAEPVNPPEPEKTPVLAKTPVKKRDPSPKAVAPASTKTDAPVSVKKEPVSKPE 865
QY 521 PAPTPKEATTPPKGAPATTLKEPATTPKAPKAPKELAPT--TTKEPTSTSDKAPATTP 579
DB 866 PSPKAPAPNSVPV-----PVPVKNPVKKM--KPPWEDEDEPTEVEKKSPSE--PEKKTPLA 918
QY 580 KGAPATTPKEPATTPKEPATTPKGTAPTLKAPAPT-----TPKKAP--KEL 627
DB 919 K-KEPEKPKD-APKVAKPRDPSKAPVPE--KEPAKVAKPRDLSPKKAIPIPAHQEA 974
QY 628 APPTTKGPTS-----TTSKP-----APTPPKT-----APTPKEPATTPKAP 670
DB 975 PPTPVKNPVKKMRPPWEDEDEPAEPVAPPEKTPVLAKKAPKAPRDP--SPKKAAPV 1031
QY 671 TPETPPTTSVSTPTTKEPTTIHKSP-----DESTPELSA-EP--TPKALENSPREPG 722
DB 1032 AAK-PDKKIEV--PPTPVKNPVKKMRPPWEDEDEPSEPVSAPEPEKTPVLAKAPKPA 1089
QY 723 V-----PTTKTPAAT-----KPEMTTAKDKTTERDLRTPPETT--TAAPK 761
DB 1090 TKPDSEAAADPVSGPTSKDKPLSKKAPVEKPKPTTDDKDLKSPKAPKEKAPAPAK 1149
QY 762 MTKETATTTKETSKTATTTQVSTTQDTTPFTKTTTLKTLAKVTTTKKTTTTE 821
DB 1150 KWKPVWDDDEPEADTTVPASKKPTEDEPADLG-----GKTKDPK----- 1193
QY 822 IMNKPEETAKPKDRTNSKATTPPKOKPTKAPKPKSTTKPKTPMRVRKPK----- 872
DB 1194 -LNKKAAPAEKPTK-----PKPKVSKPEKPTPEPKP--AAPKMKKPPMEDDPDEPE 1243
QY 873 ---TTPPKMTSTMP 885
DB 1244 ADFTMPAKPKKPDTEDP 1259

RESULT 11
Q9LIE8 PRELIMINARY; PRT; 1480 AA.
ID Q9LIE8

```

AC 09L1B8:
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE SMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCBI_TaxID=3702;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA:
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA:
 RX PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 DR EMBL: AP001306; BAB03062.1;
 DR InterPro: IPR002965; P_rich_extensn.
 DR PRINTS: PR01217; PRICHEXTENS.
 SQ SEQUENCE 1480 AA; 147153 MW; DIAOC0C79F155E732 CRC64;

Query Match 12.28; Score 800.5; DB 10; Length 1480;
 Best Local Similarity 29.68; Pred. No. 1.7e-45;

Matches 307; Conservative 54; Mismatches 505; Indels 171; Gaps 45;

QY 34 KKKPKRPVVDEAGSLDNGDFKVTTPDSTQHNKVSIS---PKITAKPINRPSL 89
 DB 117 KPRHPRKRPVVK-----PHHPRKPRTPKPHHPRKPRTPKPHHPRKPRPSV 167
 QY 90 PPNSDTSKESLYNKETVETKTTNKOTSDGKEKTSIAETOSIEKTSKDLAPT 149
 DB 168 KPPSTK-----PPTNPPSTPQRPPIHKKRPPCTPPVASSPMAATP 211
 QY 150 SIV--LAKPKKATTTKGPATTTKPEPTP--TTP--KEPATPKPEPTPTIKSAPTTP 203
 DB 212 TQMPPIATP-PIAKSPVATPIATPPATPTTTPVATPIITTPPIANPPIIMPIATP 270
 QY 204 KEPAPTT---KSAPTTPKEPATTT-----KEPATTPKEPATTTKE 244
 DB 271 PVAPPIITNPPISKRPVTTT---PTTTPPIAKPPIATPISTIPPAATPPAATPTTLP 326
 QY 245 PA-----PTTTSAPTTP-KEPATTPPKRAP--TTPKRAPTTKPEPTTPKEPAT 294
 DB 327 PAKPPVAISPIVT--PVTPIAQPPVATPPATPPVATPIATPTTSISTPISSESP 384
 QY 295 TTKEPAPTTP-KEPATPKKAPATTPKEPATTPKKEPATTT---KEPSPTPKKEPAT 349
 DB 385 VATPTPATSPKTPPAKPPVATPIAKSPVATPIATPPATPPATPIEKPPVATPTTTP 444
 QY 350 TTTKSAPTTPKEPATTTKSAPTTPKEP--SPTTTKEPATTPKEPAT--TTPKKAAPT 404
 DB 445 TAT---PPVAKPVEPTPIATPPAKPPISPPISKPPVATPPAATPTTTPVKKPPVA 501
 QY 405 TPKRAPTTKREP--APTTPKKRAPTA--KEPATTPKKEPAT-----PKLTP--T 452
 DB 502 TPIPAIPVAAAPVATPPATPIATPIATPIAKSPVATPPATPPVATPIAKPPVATPTT 561
 QY 453 TPEKLAFTTEKRAPTPPEELAPTT-PEEPPTPEEPAPTTPKAAAPNT---PKEPAPT 508
 DB 562 TPIPATPPVAKPPVATPIATPIATPIAKPPISITPPSKSVATPPAATPTTTPPAKPPVA 621
 QY 509 TPKEPAPTTPKEPATTPKKEPAT--TTPKGTAPT--LKEPATTPKKAAPKELAPTTT 563
 DB 622 TPIPATPIAKPPVATPPATPIATPIATSPVAKPPVATPPIKPPPAKPPVATPIATPPVA 681

QY 564 KEPTSTSDKRAPTPKGTAP-TTPKEPATTPKEPATTPPKGTAPTPTTIKEPATTPKRP 622
 DB 682 KPPVAT---PPTAATPIATPIATPPVATPPATPPVATPIAKPPTTI--PPTATPPVA 736
 QY 623 APKELAPTTTGGPST--TSDKRAPTPKKEPAT--TTPKEPATTPKKEPATTP-----E 673
 DB 737 MPPIATTEPTAKPPIATPIATPIAPVAKPPVATPIATPIATPIATPIAKSPVATPPATPPVA 796
 QY 674 TTPPTSEVSTPTTTKEPTTIH--KSPDESTPELSAEPPTKALENSKEPGVPTTKTPAA 731
 DB 797 TPIAKPPVATPPATPPATPPVAKPPVATPPATPPATPPATPPATPPATPPATPPATPPAA 856
 QY 732 TKPEMT--TTAKDTEEDLRTPEPTTAAPKMTKETATTTKTESKITATTTQVSTT 790
 DB 857 TPIITTPPAKPPVATPIATPP--IAKPPVATPPATPIATPIATSPVAKPPVAIPPIKTP 914
 QY 791 QDTTPFKITLTKTTLAPKVTTKKT---TTTEIMNKKPEETAKPKDRATNSKATTPKP 846
 DB 915 PAKPPVATPIATPPVAKPPVATPPATPPATPIATSPATPPVVT---PPTATSPVATPPIA 971
 QY 847 QKPTKAKRPSTSKKPKTMPVRKPKTTPPKKMTSTMPELNP-----TSRIAE--- 895
 DB 972 KPTTTP--PTAT-PPVAMPPIATP---PTAKPPVATPIANPVEKPPVATPIAKPPT 1025
 QY 896 -----AMQTTTRNQNTPNSKL--VEVNPKSEDAAGAGEPHMLLRPHVMPVETDMD 948
 DB 1026 VLPPIAKPPVETSPATPPATPPVATPPVATPP-----VAKPPVAIPPIKPPV 1070
 QY 949 YLPRVNPQIITNPMLS 965
 DB 1071 ATPPTVNPPTAMPPIVT 1087

RESULT 12
 Q41805
 ID Q41805 PRELIMINARY; PRT; 1188 AA.
 AC Q41805;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE EXTENSIN-LIKE PROTEIN PRECURSOR.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoidae; Andropogoneae; Zea.
 NC NCBI_TaxID=4577;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=B73; TISSUE=POLLEN;
 RA Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z34465; CAA84230.1;
 DR Mendel; 14346; Zeama; 2368; 14346.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 3.
 DR SMART: SM00370; LRR; 4.
 KW Signal.
 FT SIGNAL 1
 SQ SEQUENCE 1188 AA; 120980 MW; 2C77C7F8D7130149 CRC64;

Query Match 11.98; Score 784; DB 10; Length 1188;
 Best Local Similarity 29.98; Pred. No. 1.7e-44;
 Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

QY 37 PTPKPPVVDEAGSLDNGDFKVTTPDSTQHNKVSISPKTTAKPINRPSL-----PPN 92
 DB 460 PTPHSPPAD-----DIVPTTPPVGKSPRATSSPQVQPPAASTPPSLVKLSPPQ 510
 QY 93 SDTSKESLYNKETVETKTTTNKOTSDGKEKTSIAETOSIEKTSKADLAPTSKV 152

```

Db 511 APV-----SPPPVKTTSPAPIG-----SPSPPPVSVV 541
Qy 153 -----LAKTPPAETTTKPPALTTKEEPTTTKEPASTTTKEPTTTKSAPTTKEPA 207
Db 542 SPPPPVKSPPPAVGSPPPEKSPPPPAVASSPPPVKS-----PPPLVASPPPVKSP 599
Qy 208 PTTTKAPPTP-----KEPAPTTEKAPATTKEPATTTEKAPATTTSATPTT-----KEPAPT 264
Db 600 PPAVAVSPPPPVKSPPPTPVASPPPPAVASSPPPMKSPPTPTVSSPPPEKSPPPP 659
Qy 265 PKKAPATTKEPAPT-----PKE-----PPTTKAPATTKEPATTKEPAPTAKKP 315
Db 660 PAKSTPEEPTPTPTSVKSSPPPEKSLPPPLIPSPPOEKTPTPTSPK-----PSSEKP 718
Qy 316 APPTPEKAPATTKEPATTKEPSTTTKEPATTTSAPTTTKEPATTTSATTTK 375
Db 719 SP-----PKEPVSSPPTP-----KSSPPPAVSSPPTPVSSPPALAVSSPSVSSP----- 768
Qy 376 EPSPTTTKEPAPTTEKAPATTTPKKAPATTTPKEPATTTEKAPATTTKKAPAKKEPAP 435
Db 769 PPAVLSPPPAPOVKSPPPVGVSSP-----PPAKSSPPLAVSSPPQVEKTSPPAPLSSPP 827
Qy 436 TTPKETAP-----TTP-----KKLPTTPEKLAPTTPEKAP-----TTEELAPTTPEPT 482
Db 828 LAKSSPPHVAVSSPPPVVKSPPPAVSSPPLTPKASPPAHVSSPEVVKSTP-----PA 885
Qy 483 PTPPEEAPPTTKAAPTKE-----PAPTTEKAPATTTPKEPATTTPKETATTTKGAPT 541
Db 886 PTVVISP-----PSEPKSSPPTVSLPPVKSPPPAVSSP-----PMTPVSSPPVVSSPPT 943
Qy 542 LKE-----PA-----PTTPKK-----PAKELAPTTTKE-----PTSTSDKAPPTTGTAPT 587
Db 944 VKSSPPPAVSSPPPAVSSPPPAVNLPPPEVKSPPPTPVSSPPA-----PKSSPPAP 1000
Qy 588 -KEPAPTTEKAPATTTPKGTAPTTTKEPAPTTPKKAPKELAPTTTGTSTSDKAPT 646
Db 1001 MSSPPPEVKSPPPAVSSPPPVKSPPPAVSSPP-----PVKSSPPPAVSSPPPV 1057
Qy 647 -TPEKAPTTPKEPAPTTPKKAPATTTPETPTTSEVSTPTTKEPTTIHKSPPESPEL 705
Db 1058 KSPPPAVSSPPPVKSPPPAVS-----SPPPVKSPPPPAVSSPPPLKSSPPPAV 1115
Qy 706 SAEPPTPALENSPKPGVPTTKPATKPEMTTAKDKTTERDLRTTPT 755
Db 1116 SPPAP-----VKPPLP-----PPAVSSPPVTPAPKKKEGSLPPAES 1158

RESULT 13
P70670 PRELIMINARY; PRT; 2187 AA.
AC P70670:
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
MUSCLE-SPECIFIC FORM GP220).
NCNA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_taxid=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=66312450; PubMed=6698236;
RA YOCOV W.V., St-Arnaud R.;
RT "Differential splicing in of a proline-rich exon converts alphanac
RL Genes Dev. 10:1763-1772(1996).";
DR EMBL; U48364; AAB18734.1; -;
DR EMBL; U48364; AAB18732.1; -;
DR MGD; MGI:106095; Nac.
DR InterPro; IPR002715; NAC.
DR InterPro; IPR003037; TS-N.

```

```

DR Pfam; PF01849; NAC; 1.
DR Pfam; PF02094; TS-N; 1.
SQ SEQUENCE 2187 AA; 220599 MW; 003646AA864DEBFD CRC64;

Query Match 11.9%; Score 778.5; DB 11; Length 2187;
Best Local Similarity 27.9%; Pred. No. 7.8e-44;
Matches 292; Conservative 119; Mismatches 420; Indels 215; Gaps 48;

Qy 22 VSSVADNKKNRKKKPTKPPVVDAGSLDGDCKVITPTSTQOHK-----VTSPPK 76
Db 803 ISAVOSPKVDPTMSDVTPTSPKTS-----TAVPKDTSATLSKVAVTSLSP 853
Qy 77 ITTAKPINDPSPNDKSTSL-----TVNKET-----TVEKETTITNKOTSDGKEX 128
Db 854 KAPVAPSNATIVPTETIPISLKNALAAATPKETLATSIPKVTSPQKPKSVSLKGAPA 913
Qy 129 TTSAKETOSIEKTSADLAPTSKVLAKPPKAETTTGALTTPEPTTTKEPASTTP 188
Db 914 MTSKAT-----EIAASKDVSPSQ-----EPKEVPLQHV-----PPTSPPKSPVSDTSLGALTSP 965
Qy 189 KEPTPTTISAPPTPE-----PAPTTTSAPTTPK-----BPAPTTEKAPATTPEKA 238
Db 966 PKGPPATLLETPTPKSPKPAASKTTPATPSPEGVAVPLEIPPSKAPATAKES 1025
Qy 239 PTTKEPAPT-----TKSAPT-----TKEPAPTTPKK-----PAPTTPK 274
Db 1026 AITSSKAPKRTANSKIEIPKGVAVPLEISLPLEKTSKATPOEKSASPSKRSPTAGPK 1085
Qy 275 EPAPTTPKEPTTPTEKAPATTKEPAPTTPKEAPAPKAPATT-----PKEAP-----TTP 328
Db 1086 E-----TPPGVAVPPEISLPKETPDONAAPEISLASSOSKRSKTSVPEETPGGVAMP 1142
Qy 329 KEAPATTTEKESPTTPK-----EPATTTSAPTTTKEPAPTTP-----KSAPTTPK----- 375
Db 1143 LEIPSAPOKAPKTAVKQIPTPEDATYTLAAGSLSKKASKTAAPKAPAVSGVAVS 1202
Qy 376 -EPSPTTKEPAPTTEKAPATTTPKKAPATT-----PKE-----PAPTTPKEAP----- 419
Db 1203 GEISPSPKKTSAAPEKNSATLPPKRSKTAAPKETAPEKATSSGVAVSEISPSPTPA 1262
Qy 420 -----TTTKAPATYKAPETTKETATTTKKLPTTPEKLAPTTPEK-----PAPTPEE 472
Db 1263 SKGVAVTLTPKGAPEALAE-----SPASPKVPKTAAPPEETSTP-----SPQIKKVAQPK 1315
Qy 473 LAPTTPEEPTPT-----PEEAPATTPKAAPTTPKEPAPTTPKE-----PAPTTPKEAP 523
Db 1316 ASATPEKKTPTKTAIVKETSAPSEGVAVPLEIPSPKAPKTAAREMTPAS-----PEGAT 1373
Qy 524 TTPKETAPTTKGTAPTTTKEPAPT-----KKEPAPKELAPTT 563
Db 1374 TAPVQIPSPKSKSKKAGSKE-----TPTPSPEGVTAAPLEIPSSKTKSKMASPKETLVPS 1432
Qy 564 KEPT-----TSDKAPATTTPKGTAPTTKEPAPTTPKE-----PAPTTPKGTAPTTKEPAP 616
Db 1433 SKTISQTVGKEKTSLEGATVPLEIPSSHKAPEYDVKOVLTPSPK-----DAPTTLAE-----SP 1490
Qy 617 TTPKAPKELAPTTTGTSTSDKAPATTTPKETATTTKEPAPTTPKKAPATTPEP----- 675
Db 1491 SSPKK-----APKTAAPSER-----VTVPEKPA-----TPQKASGTTASKVPAVATQVAVSSRETVP 1547
Qy 676 -----PPTTSEVSTPTTK-----EPPTTIHKSPEESTPELSAEPPTPALENSPRE-----PGVP 724
Db 1548 TPAPVPVKNSSHKTSKTIELKEAPATLPPSPKSPKISSKAPRT-----SAPKEFASP 1605
Qy 725 TTKPATKPEMTTAKDKTTERDLRTTPETTTAAKMTKETATTTKEKTSKATTTATTTQ 784
Db 1606 SIK-----PVT-----SLAQTAPPSLOKASTIIPKENLAAPV-----LPVSSKSPAAPARAASL 1658
Qy 785 VTSTTODTTPKITTLLKTTTLAPKVTTKTITTEIMKPEETAKPKORATNSKATTP 844
Db 1659 SPATAAPQATAPKEATTITPSCKAAATETPIETSTAPSLGAPKETSE-----TSVSKVIMS 1714

```

QY 845 KPOKPTKAPKPTSTKPKTPRVKPKTPTTPRKATSMPELNPTRIAEMLQTTTRP 904
 Db 1715 SP-----PKKSSSKRASTLP-----ATTPLSLKASVLS-----P 1745
 QY 905 NOTPNSKLVENPKSEDAGAGAGETP 930
 Db 1746 TATSGKDSHISPV-S-DACSTGTTP 1770
 RESULT 14
 QYVEL9 PRELIMINARY; PRT; 2112 AA.
 AC QYVEL9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CG4090 PROTEIN.
 GN CG4090.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borokva D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iregyan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003716; AAF55402.1;
 DR Flybase: Fgn0038492; CG4090.
 DR InterPro: IPR002357; Chitin_binding.
 DR Pfam: PF01607; Chitin_bind_2; 10.
 DR SMART: SM00494; ChitBD; 11.
 DR PROSITE: PS00022; EGF_1; 1.
 SQ SEQUENCE 2112 AA; 219547 MW; B91018E551A5D36 CRC64;

Query Match 11.3%; Score 739; DB 5; Length 2112;
 Best Local Similarity 27.2%; Pred. No. 3.3e-41;
 Matches 281; Conservative 98; Mismatches 339; Indels 316; Gaps 44;
 QY 28 NKKRRTKPKPPKPPVVDAGSGJLNDGDFKVTTPDTSTQHNKVSTPKTTAKPIMNRP 87
 Db 1023 NHPDQIQTKPLCKKRVYSGGSSSN-----STNSSSSSNNSSSSSSSSSSSS----- 1070
 QY 88 SLPPNSDPTKETSLLVKKETV---ETKETTITKOTS---TDGKKTSAKTOSI-- 138
 Db 1071 SSSNSGSSSSNTGSSNSGASSSGCGSSNQGSSSSGSSSSGSSGSSSTSSSSSS 1130
 QY 139 ---EKTSANDIAPTSKVLAKPPTKAAETTKGP----- 167
 Db 1131 NNNNGSSSSSSSSSSSSSSSSSSSSSKPNPSECKVNGQFIDRSDCAKFRVCDNDRGFNWPF 1190
 QY 168 -----ALTTKREPTTPPKKPASTTPKEPTTTKSAPTT 202
 Db 1191 SCGGYWDQMOMQACNHAMAVKECGIAPPTSTPTSR-PTASTSRPSDQSTSRPTG 1249
 QY 203 KKEPAPTTKSAPTTPKKEPAPTTTTPKKEPAPTT----- 241
 Db 1250 P-----PTTARVYTAAPTSTSSPTTASSSQTTSPTVQAQNTGCKRSEGFMAIDPNKCSFYR 1305
 QY 242 -----TKEP---APT-----TKSAPTTPKKEPAPTTKKEPAPTT 272
 Db 1306 CVRNKKGFTSIPQCGAGYWDQDLOTCNHNFNCSSTGTESTPPKPC--EPATNGTTA 1363
 QY 273 KKEPAPTTKEPPTTPKKEPAPTTKKEPAPTTKKEPAPTTAKKAPTTKKEPAPTTKKEPA 332
 Db 1364 TSSSTTTP--PPTTDLPTSTTGAP-PTTTELPPT-----TTDLPPTTTTLRLP 1413
 QY 333 PTTKREPTTPKKEPAPTTKSAPTTPKKEPAPTTKSAPTTPK--EPSPPTTKKEPAP-- 387
 Db 1414 TTTTSLPPTTTGAPPTTGAQPTTTTSSSETSVITVTSPESTQPPSTTKMLKPLAG 1473
 QY 388 -----TTK-----TTK-----EPAP---TT 397
 Db 1474 TECTGEGYADPEDCKRYKNCINAGASYRKYNTFCRKGTMNEVOYTCDEVENIPKCSKL 1533
 QY 398 PKRPAPTTTP---KEPAPTTPKKE--PAPTTTKKAPAPKKEPAPTTKETAAPTTKLLPT 452
 Db 1534 PAPTITTPKEEKDGSSTTQSTDEPTVTKI-TPKE-ESTKPKPKPTQVPEKPT 1591
 QY 453 TPKEAPTTPEKAPTTPEELAPTTPEEPTTPTEEPAPTTPKAAAPNTPKEPAPTTKPE 512
 Db 1592 TTEE--PEKPKPT--TTEVPKPTTTEEPKPKPT--TTEVPKPTTTEEPKPKPK 1647
 QY 513 PAPTTP--KKEPAPTTPKETA-----PTTPK--GAPTTTKEPAPTTKPKAPKELAPTTTK 564
 Db 1648 PTTTEPQKPTTEEPPTTSLPGNPTTTSVPGNPTTTPPIVHTTSTGKYK---PTTGG 1704
 QY 565 EPTSTSDKRAPPTPKGAPTTAPTTKKEPAPTTPKK--PAPTTPKGAPTTTKEPAPTTKPK-- 621
 Db 1705 EPITTTT-LPSTTTDAIOETTSKKKEPTTTTSPSSSTPEGSVTTLOPEPQPNWCSSE 1763
 QY 622 ---PAKELA-----PTTKGPTSTSKPAPTTPKETAAPTTPKKEPAPTTKPKAPPTP 672
 Db 1764 GFEPDEDSRYRCVDAKNGKYVAFKCGKGTWMDSTECNVADQVSGN----- 1816
 QY 673 ETPPTTSVSTPTTTTKEPTTIHKSDESTPELSAEPPTKALENSKPEKGVTTTTPAAT 732
 Db 1817 ---SSGQTTTPGTTTEGTTSTSSGK---ETTSKABEN-----TTTWA-- 1857
 QY 733 KPEMTTAKDKTTERDLPTTPETTTAAPKMTKEFATTTTKEKTSKITATTQVSTTQD 792
 Db 1858 -EPTTT-----SSPETTT--VASETTTTTSGT-----TTATPPTTTKP 1895
 QY 793 TTPFKITTLKTTLAPKVTYTTTKKTIYTTTEIMNKPETAKPKDRATNSKATPPKPKPTKA 852
 Db 1896 PKP-----ETTTIAGEETSTKSPTTT-----SAPSTNTSA 1928

QY 853 PKKPTSTKKPKMP 866
Db 1929 P-----CPETGP 1935

RESULT 15

Q9XDH2 PRELIMINARY; PRT; 763 AA.
AC Q9XDH2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
OC PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA Espitia C., Lacleite J.L., Mondragon-Palomino M., Martens A.,
RA Zhang Y., Moreno C., Singh M.;
RT "Cloning and characterization of a new member of the PGHS family that
RT is a useful marker of polymorphism in Mycobacterium tuberculosis."
RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF071081; AAD41594.1;
DR InterPro: IPR002851; Actophin.
DR InterPro: IPR002965; P_rich_extensn.
DR PRINTS: PRO1222; ATROPHIN.
DR PRINTS: PRO1217; PRICHEXTENSIN.
DR PRINTS: PRO1218; PSTEXTENSIN.
SQ SEQUENCE 763 AA; 75034 MW; 39168EC45A5916F8 CRC64;

Query Match 10.9%; Score 715; DB 2; Length 763;
Best Local Similarity 28.2%; Pred. No. 4,6e-40;
Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

QY 156 PTPKAEITTKGPAALTTPKEPTTPKEBASTTPKEPTPTTKSAPTTPKEBAPTTPKSA 215
Db 3 PVP-----APRALAPLPAPAPAEKSKPPPPAP-----PAPCWMIVSAP 46
QY 216 TTPKEBAPTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKE 275
Db 47 PCF--PAP--PAPKPKSKAPFPVPPAPARELAPLP--PAP-----PEAPRE 90
QY 276 PAPTPKEPTPTT--PKBAPTTPKEBAPTTPK--EPAPTAPK--KPAPTTPKEBAPTTPK 329
Db 91 SRPALPCPPPPVYIPDPPEBAAPVPAPANSPPFPFPAPKFAVPVP--PVPNSPP 148
QY 330 EP--APTTKEPSPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEPSPTTPKEBAP 387
Db 149 FPPPPRALNPPAP--PAPPLANSPLPAPAPTPAGT--PPAAMPVPPAPAPKSKPA 201
QY 388 TTPKEBAPTTPKPAPTTPKEBAPTTPKSAPTTPKEBAPTTPKSAPTTPKEBAPTTPKETAP 443
Db 202 SPPRPAP--PMATGWEPFLPPVPDPISKETPPAPAPPTIPAPVPPIPVPLP 256
QY 444 TTPKKLTTPPEKLAP-----TTPKEBAPTTPPELAPTTPEEPTPTTPPEBAPTTP-- 493
Db 257 PVPKTIAPAP--APVAVAAVLVABCPPLPLPNNHPPAPAPAVPVGVLAPLPLNSHP 313
QY 494 ---PKAAAPMTPKEPAP----- 507
Db 314 PAPPSAPVPGVPLAPLPSGRPVSVKGSFTTLSTFCRCVSGEVLALNPSRPSRSLP 373
QY 508 --TTPKEBAPTTPKEBAPTTPKETA--PTTPKETA--PTTPKEBAPTTP--KKPAPE 557
Db 374 TTTTPALPAPIPPLPPLPINTAVPPIPLPVVALAPLPLPPLAPLPSGVPAP-- 431
QY 558 LAPTTPKEPTSTSDKRAPPTPKGTAPTPKEPA-----PTTPKEBAPTTPKGTAPPTL 611

Db 432 --PIPPGKFWITPLAPAPPEPK--TVPVLPPGBCPSPSEKPNPPAPPEPEPKSSPALPP 488
QY 612 KEBAPTTP--KKPAPKELAPTTKGP--TSTTSOKPAPTTPKET--APTTPKEBAPTTP 664
Db 489 APPAPSMPSAVRPDPSPPIPPAPAPAPRASMPALPPAPSPPATRLCPPLPSPPAVNSP 548
QY 665 KKPAPTTPETPTTPSEVSTPTTPKEPTTIHKSP--DESTPELSAETPKA--LENSP 718
Db 549 --PAPAPAPPPKLLS--ANPPCPVPAPNRPAPAPAPAPLAPAPDPPTPPVANSP 604
QY 719 KEGVPTTKTPPAATKPMATTTAKDTERDLRTPETTAARPKMETATTTETTESKI 778
Db 605 --PAPAPAPPSALPVPNPPA-----PTTPAAPK-----SRPAL 637
QY 779 TATTOVSTTTODTTPFKITTLTKTTLAPKVTTTKTITTEIMNKPEETAKPKDQATN 838
Db 638 PAAPAPAPAPVPRATP-----PAPAPAPAPN 665
QY 839 SKATTPKQKP--TKAPKKPTSTKKPKTPPVKPKKTTPTPRKMTSTW----DELNP 889
Db 666 SMALPPAPDPDPLLATPPAPAPPLPMSPPAPPLPPAPADPPAPDLITNQPPSPLAP 725
QY 890 TSRLAEMLQTTTRPNQTPNSKIVEVNPKSEDAAGAGET 929
Db 726 VPGAPLAPLPIINGRPVARKNSLI-----GSSSGOT 756

Search completed: April 26, 2002, 16:28:53
Job time: 548 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:29:32 ; Search time 210.7 Seconds
(Without alignments)
948.306 Million cell updates/sec

Title: A66
Perfect score: 7294
Sequence: 1 MAWKPLPYLLLSVFYIQ.....ARAITRSGQTLKWMYNCNP 1366

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SP:REMBL_17:*
2: SP:archaea:*
3: SP:bacteria:*
4: SP:fungi:*
5: SP:human:*
6: SP:invertebrate:*
7: SP:mhc:*
8: SP:organelle:*
9: SP:phage:*
10: SP:plant:*
11: SP:rodent:*
12: SP:virus:*
13: SP:vertebrate:*
14: SP:unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7265	99.6	1404	4	Q92954	Q92954 homo sapien
2	7236	99.2	1404	4	Q9BX49	Q9BX49 homo sapien
3	3759	51.5	1054	11	Q9JM99	Q9JM99 mus musculus
4	1713.5	23.5	401	6	077765	077765 bos taurus
5	982.5	13.5	1079	5	Q9N4S7	Q9N4S7 caenorhabd1
6	946.5	13.0	1049	5	Q9I7S1	Q9I7S1 drosophila
7	891.5	12.2	1795	5	076894	076894 drosophila
8	863.5	11.8	1489	10	Q96449	Q96449 phytophthor
9	846	11.6	1274	5	Q20007	Q20007 caenorhabd1
10	817	11.2	1315	10	Q9SPM0	Q9SPM0 zea mays (m
11	814	11.2	1480	10	Q9LIE8	Q9LIE8 arabidopsis
12	812	11.1	2187	11	P70670	P70670 mus musculus
13	784	10.7	1188	10	Q41805	Q41805 zea mays (m
14	781	10.7	2112	5	Q9VEL9	Q9VEL9 drosophila
15	715	9.8	763	2	Q9XDH2	Q9XDH2 mycobacteri
16	707.5	9.7	555	5	Q9FPQ6	Q9FPQ6 chlamydomon
17	706.5	9.7	6677	5	Q9N435	Q9N435 caenorhabd1
18	703.5	9.6	4880	11	Q9JUL1	Q9JUL1 rattus norv
19	703.5	9.6	5085	11	Q9JkS6	Q9JkS6 rattus norv

20	695	9.5	1151	13	057580	057580 gallus gall
21	680	9.3	1229	5	094185	094185 caenorhabd
22	667.5	9.2	2284	5	09VPG1	09VPG1 drosophila
23	665	9.1	4833	11	09QYX6	09QYX6 mus musculu
24	665	9.1	5038	11	09QYX7	09QYX7 mus musculu
25	664	9.1	2089	4	014676	014676 homo sapien
26	661.5	9.1	7962	4	010465	010465 homo sapien
27	660.5	9.1	2768	5	09VC00	09VC00 drosophila
28	653.5	9.0	1514	5	09GUM7	09GUM7 leishmania
29	649	8.9	3570	5	09G552	09G552 homo sapien
30	647.5	8.9	3507	5	023587	023587 caenorhabd
31	643.5	8.8	990	13	091803	091803 xenopus lae
32	636	8.7	6632	5	017362	017362 caenorhabd
33	633	8.7	489	10	041707	041707 vigna ungu
34	632	8.7	761	10	09Z010	09Z010 arabidopsis
35	631	8.7	2344	5	09N3X8	09N3X8 caenorhabd
36	629	8.6	1642	5	001761	001761 caenorhabd
37	628.5	8.6	1612	5	09VYQ2	09VYQ2 drosophila
38	627.5	8.6	971	5	09XVS4	09XVS4 caenorhabd
39	622.5	8.5	839	2	09RXS7	09RXS7 delnoccocus
40	607.5	8.3	801	5	023635	023635 caenorhabd
41	607.5	8.3	924	12	099307	099307 epstein-barr
42	605	8.3	379	5	027929	027929 drosophila
43	600.5	8.2	1893	5	09NKC9	09NKC9 drosophila
44	598.5	8.2	409	10	09SBM1	09SBM1 volvox cart
45	592.5	8.1	2232	5	P91365	P91365 caenorhabd

ALIGNMENTS

RESULT	ID	SEQUENCE FROM N.A.	PRELIMINARY:	PRT:	1404 AA.
092954	092954	01-FEB-1997 (TREMBLrel. 02, Created)			
AC	092954	01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	MEGAKARYOCYTE STIMULATING FACTOR.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,				
RA	Kellerher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,				
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,				
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;				
RA	Blood 78:279-279(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,				
RA	Fitzgerald M., Scalitro J., Kellerher K., Preissner K., Kriz R.,				
RA	Jacobs K., Turner K.,				
RL	(in) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,				
RL	Mosher D.F. (eds.);				
RL	Biology of vitronectins and their receptors.. pp.45-52,				
RL	Elsevier Science Publishers B.V. (1993).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,				
RA	Kellerher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,				
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,				
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;				
RA	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.				
EMBL	U70136; AAB09089.1; -				
DR	InterPro: IPR000585; Hemopexin.				
DR	InterPro: IPR001212; Somatomedin-B.				
DR	InterPro: IPR002400; GF_cysknot.				
DR	Pfam: PF00045; hemopexin; 2.				
DR	Pfam: PF01033; Somatomedin-B; 2.				
DR	PRINTS; PR00436; GFCYSKNOT.				


```

OY 143 KISSKNSAANRELQKLVKDNKKNRKKPKPPVVDAGSGLDNGDFKVTTPDST 202
    |||
Db 181 KISSKNSAANRELQKLVKDNKKNRKKPKPPVVDAGSGLDNGDFKVTTPDST 240
OY 203 TOHNKYSTPKITTAKPINRPSLPPNSDTSKETSITVKKETTVETKETTNNKOSTD 262
    |||
Db 241 TOHNKYSTPKITTAKPINRPSLPPNSDTSKETSITVKKETTVETKETTNNKOSTD 300
OY 263 KKKTSKETSQSIKTSANDLAPTSKVLAKPTPKAETTTKGPAITPKPEPTTTPKEKAS 322
    |||
Db 301 KKKTSKETSQSIKTSANDLAPTSKVLAKPTPKAETTTKGPAITPKPEPTTTPKEKAS 360
OY 323 TTPKEPTPTTISAPTTKEBPATTTKSAPTTKEBPATTTKEBPATTTKEBPATTTKEP 382
    |||
Db 361 TTPKEPTPTTISAPTTKEBPATTTKSAPTTKEBPATTTKEBPATTTKEBPATTTKEP 420
OY 383 APPTTKSAPTTKEBPATTTKPKAPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTK 442
    |||
Db 421 APPTTKSAPTTKEBPATTTKPKAPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTK 480
OY 443 BPAPTPAKKPAATTTKEBPATTTKESPTTKEBPATTTKSAPTTTKEBPATTTKEBPATTT 502
    |||
Db 481 BPAPTPAKKPAATTTKEBPATTTKESPTTKEBPATTTKSAPTTTKEBPATTTKEBPATTT 540
OY 503 TTKSAPTTKEBPATTTKEBPATTTKEBPATTTKPKAPATTTKEBPATTTKEBPATTTKPK 562
    |||
Db 541 TTKSAPTTKEBPATTTKEBPATTTKEBPATTTKPKAPATTTKEBPATTTKEBPATTTKPK 600
OY 563 APAPKBPATTTKEBPATTTKPKLPTTPEKLAPTTPEKLAPTTPEKLAPTTPEKLAPTTPEK 622
    |||
Db 601 APPTTKEBPATTTKEBPATTTKPKLPTTPEKLAPTTPEKLAPTTPEKLAPTTPEKLAPTT 660
OY 623 PEEBPATTPKAAAPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 682
    |||
Db 661 PEEBPATTPKAAAPTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 720
OY 683 APPTPKKPAKELAPTTTKEPTSTSDKPAATTTKGAATTPKGAATTPKGAATTPKGAATTTKEP 742
    |||
Db 721 APPTPKKPAKELAPTTTKEPTSTSDKPAATTTKGAATTPKGAATTPKGAATTPKGAATTTKEP 780
OY 743 TAPPTLKBPATTTKPKAPKELAPTTTGGPTSTSDKPAATTPKEPATTPKBPATTTK 802
    |||
Db 781 TAPPTLKBPATTTKPKAPKELAPTTTGGPTSTSDKPAATTPKEPATTPKBPATTTK 840
OY 803 KPAPTTPEPTPTTSEVSTPTTKEPTTIKSPDESPBELSAPTPKALNSKKEGCVPT 862
    |||
Db 841 KPAPTTPEPTPTTSEVSTPTTKEPTTIKSPDESPBELSAPTPKALNSKKEGCVPT 900
OY 863 TKTPAATKPEKTTTAKDKTERDLRTTPETTTAAPKWKETATTTTEKTTESKITTATTTTOY 922
    |||
Db 901 TKTPAATKPEKTTTAKDKTERDLRTTPETTTAAPKWKETATTTTEKTTESKITTATTTTOY 960
OY 923 TSTTTOOTPEKTTTLLKTTTLLAKVTTTAKTTTTELMNKPEETAKPKONATNSKATTPK 982
    |||
Db 961 TSTTTOOTPEKTTTLLKTTTLLAKVTTTAKTTTTELMNKPEETAKPKONATNSKATTPK 1020
OY 983 POKPTPAKPKPTSTKPKTPVRAKPKTPTPKKMTSTMBELNPTSRIEAMQOTTTRN 1042
    |||
Db 1021 POKPTPAKPKPTSTKPKTPVRAKPKTPTPKKMTSTMBELNPTSRIEAMQOTTTRN 1080
OY 1043 QTPNSLIVENPKSEDAAGAGETPHMLLRPHVEMPEVTPDMDYLRPVDNOGIINPMLS 1102
    |||
Db 1081 QTPNSLIVENPKSEDAAGAGETPHMLLRPHVEMPEVTPDMDYLRPVDNOGIINPMLS 1140
OY 1103 DETNINCKGRVVDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRTTEVAGISPIDYVFT 1162
    |||
Db 1141 DETNINCKGRVVDGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRTTEVAGISPIDYVFT 1200
OY 1163 KRCCEKTEFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIVALASTAKKKNPESY 1222
    |||
Db 1201 KRCCEKTEFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIVALASTAKKKNPESY 1260
OY 1223 FFKRGSIOQYIYKQEPVQKCGRRPALNVPYGEHTOVRRRRERBAIGSOTHTIRIOY 1282

```

```

Db 1261 FFKRGSIOQYIYKQEPVQKCGRRPALNVPYGETTOVRRRRERBAIGSOTHTIRIOY 1320
OY 1283 SPARLAVODKGVNLHNEKVSITLMRGLPNVYTSALSLEINRKPDPGYDYAFSKDOYVINDY 1342
    |||
Db 1321 SPARLAVODKGVNLHNEKVSITLMRGLPNVYTSALSLEINRKPDPGYDYAFSKDOYVINDY 1380
OY 1343 PSRTARAATTRSGOTLSKRWYNCP 1366
    |||
Db 1381 PSRTARAATTRSGOTLSKRWYNCP 1404

RESULT 3
O9JM99 PRELIMINARY; PRT: 1054 AA.
ID O9JM99;
AC O9JM99;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE mRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN PRG4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikegawa S., Nakamura Y.;
RT "a novel mouse gene highly homologous to a human gene encoding
RT megakaryocyte stimulating factor precursor and cartilage superficial
RT zone protein."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB034730; BAA92310.1; -.
DR MGI: 1891344; Prg4.
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR01212; Somatomedin_B.
DR Pfam: PF01033; Somatomedin_B.
DR PRINTS: PR01217; PRICHEXTENS.
DR PRINTS: PR00022; SOMATOMEDINB.
DR SMART: SM00120; HX; 2.
DR SMART: SM00201; SO; 2.
DR PROSITE: PS00524; HEMOPEXIN; UNKNOWN_1.
DR PROSITE: PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1054 AA; 115991 MW; 4fC64BFA42283235 CRC64;

Query Match 51.5%; Score 3759; DB 11; Length 1054;
Best Local Similarity 55.0%; Pred. No. 8,66-237;
Matches 779; Conservative 56; Mismatches 169; Indels 412; Gaps 26;

OY 1 MAMKTLPIYLILLLSFYIQOVSODLSSCAGRGEGYSRDATCNCDCVNCQHYMECCPDF 60
    |||
Db 1 MAMKTLPIYCLLSLPLVLIQOVSODLSSCAGRGEGYSRDATCNCDCVNCQHYMECCPDF 60
OY 61 KRVCY-----AAEVNPPSPSSSKAP 82
    |||
Db 61 KRVCSPELSCGRCPESFARGREDCDCSQOYQKCCADVDYFCEEVHNSTSPSSKAP 119
OY 83 PPSGASQITKSTTRSPKPPKKTKKAYIESEITTEHSVSENOSSSSSSSSSTIW 142
    |||
Db 120 TPAGASDTIKSTTRSPKSPPT-TRTIKVESEELTEHSSENGE--SSSSSSSSSTIR 176
OY 143 KISSKNSAANRELQKLVKDNKKNRKKPKPPVVDAGSGLDNGDFKVT--TPDT 200
    |||
Db 177 KISSKNS-ANRELQKNPNVANDKNKPKKPNPEPAVDEAGSGLDNGDFKLTTPPDP 235
OY 201 STOHNKYSTPKITTAKPINRPSLPPNSDTSKETSITVKKETTVETKETTNNKOST 260
    |||
Db 236 PTPHRSKATSPKTTAKPVTPKPSLAPNSETSSEASLASNKEETVETKETTATNKOSSA 295
OY 261 DKKETTSKETSQSIKTSANDLAPTSKVLAKPTPKAETTTKGPAITTPKEPTTTPKEP 320

```

```

Db 296 -SKKTTSVKETSASAKTSDKV-----EP 319
QY 321 ASPTPEPTTTTTSAPTTKPEAPTTTTSAPTTKPEAPTTTKEAPTTTKEAPTTT 380
Db 320 TSTTTPK-----NSAPTTTKRPV--TTTKESKLP-----LPQDEPTTAK 357
QY 381 EPAATTTKSAPTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTT 440
Db 358 EPPPTTKKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTT 417
QY 441 PKRPAPAPKRPAPTTKPEAPTTTKEAPTTTKEPPTTKEPPTTKEPPTTKEPPTTKEP 500
Db 418 PKRGPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTT 461
QY 501 PTTTSAPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTT 560
Db 462 -----EPTTPEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTT 512
QY 561 KPAPTAKEAPTTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEP 620
Db 513 -----512
QY 621 TTPEAPATTPKAAAPNTPKPEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAP 680
Db 513 -----TPKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEP 539
QY 681 EPAATTTKRPAPKELAPTTTKEPPTTSDKAPATTPKGTATTTTKEAPTTTKEAPTTT 740
Db 540 EPEPTTPKRP-----EPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEP 570
QY 741 KGTAATTTKEAPTTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTT 800
Db 571 -----KEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTT 609
QY 801 PKRAPATTPPEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEPPTTKEP 860
Db 610 PKKPEPTT-----617
QY 861 PTTKPAATKPEMTTAKDCTTERDLRTPTPTTAAAPKMTKETATTEKTTESKITATTT 920
Db 618 -----TSPT-----622
QY 921 OVTSTTODTTPKTTTTLKTTLAKVTTTTLKTTLTTELMKPEELAKPKDRACTNSKAT- 979
Db 623 -----TTTKATTLAKVAPAE-----EIQNKPEPTTTPASESDSKTTL 662
QY 980 -----TPKP--OKPTKAPKPTSTKPKPTMPRVKPKPTTTPPKMTSTMBELNFTSRIA 1031
Db 663 KFOKPTKAPKPTKPTKAPKPTSTKPKPT--PKTKRPKTTTAPLKTTSATPELANTP--L 719
QY 1032 EAMLOTTTRPNOTPNKSKLEVNPKSEDAAGAGETPHMLLRHVMPETVTPMDVLPV 1091
Db 720 EVMALPTTTPKOTPNPETLEVNPDEHDADGGEKRP--LPGPVLPFPAIPETDLLAGRL 778
QY 1092 NCGIINPMALSOETNLCNKPVDGLTTLNGLTLVAFRGHYFWMLSFSPSPSPARRITEVW 1151
Db 779 NRGININPMPSDETNLNCKNKPVDGLTTLNGLTLVAFRGHYFWMLEFRPSPSPRRITEVW 838
QY 1152 GISPSPIDVETTRCNCEGKTEFFFKDSQYWRFTNDIKDAGYPKLIFKFGGLTGOIYAALST 1211
Db 839 GISPSPIDVETTRCNCEGKTEFFFKDSQYWRFTNDVYDGPYKOIVKFGGLTGOIYAALST 898
QY 1212 AKKKNMPESEYFFKRGSGTQOYTYKQEPVOKCGRRPALNTYVYGGMTVRRRRFEPAIG 1271
Db 899 AKYKDRPESYFFKRGSGTQOYTYKQEPVOKCGRRPALNTYVYGGMTVRRRRFEPAIG 958
QY 1272 PSQTHIRIOYS--PARLAYODKGVLHNEVKSILMRGLPNVYTSALSLPNIRKPDGYDY 1330
Db 959 PFOTHIRIHYSPMAKVSYODKGVFLHNEVKSILMRGLPNVYTSALSLPNIRKPDGYDY 1018
QY 1331 AFSKDOYNNIDVPSRTARAITTRSGOTLSKIWNCP 1366

```

```

Db 1019 AFSKDOYNNIDVPSRTARAITTRSGOTLSKIWNCP 1054
RESULT 4
ID 077765 PRELIMINARY: PRT: 401 AA.
AC 077765;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SUPERFICIAL ZONE PROTEIN (FRAGMENT).
OS Bos taurus (Bovine), Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-ARTICULAR CARTILAGE;
RA Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,
RA Aydelotte M.B.;
RT "Immunodetection and partial cDNA sequence of the proteoglycan,
RT Superficial zone protein, synthesized by cells lining synovial
RT joints.";
RL J. Orthop. Res. 0:0-0(1998).
DR EMBL: AF056218; AAD13404.1; -.
DR InterPro: IPR000585; Hemopexin.
DR Pfam: PF00045; hemopexin; 2.
DR PROSITE: PS00024; HEMOPEXIN; UNKNOWN_1.
DR SMART: SM00120; HX; 2.
FT NON_TER 1
SQ SEQUENCE 401 AA; 44952 MW; 86147CC9AFB73D7 CRC64;
Query Match 23.5%; Score 1713.5; DB 6; Length 401;
Best Local Similarity 80.2%; Pred. No. 4.8e-104;
Matches 320; Conservative 24; Mismatches 52; Indels 3; Gaps 3;
QY 969 PKDRATNSKATTPKPKPTKAPKPTSTKPKPTMPRVKPKPTTTPRK--MTSTMBELNPT 1027
Db 5 PKGRATNSQVTTTPKPKPTKAPKPTSTKPKPT--PVRKPKTTTPPKTTTSAMDEPPT 63
QY 1028 SRIAEAMLOTTTRPNOTPNKSKLEVNPKSEDAAGAGETPHMLLRHVMPETVTPMDVLP 1087
Db 64 S-LEAMLOTTTRTPPNSEIIDVNSENEDGAAGSEKPHMIFRPVLPVITPOTETII 122
QY 1088 PRVNOGIIINPMALSDETNLCNKPVDGLTTLNGLTLVAFRGHYFWMLSFSPSPSPARRI 1147
Db 123 VRGSGQFNGINPMPSDETNLNCKNRPVDGLTTLNGLTLVAFRGHYFWMLEPFPPTPRRI 182
QY 1148 TEVWGISPIDVETTRCNCEGKTEFFFKDSQYWRFTNDIKDAGYPKLIFKFGGLTGOIYA 1207
Db 183 TEVWGISPIDVETTRCNCEGKTEFFFKDSQYWRFTNDIKDAGYPKLIFKFGGLTGOIYA 242
QY 1208 ALSTAKYKKNMPESEYFFKRGSGTQOYTYKQEPVOKCGRRPALNTYVYGGMTVRRRRFE 1267
Db 243 ALSTAKYKKNMPESEYFFKRGSGTQOYTYKQEPVOKCGRRPALNTYVYGGMTVRRRRFE 302
QY 1268 RAISPQTHIRIOYS--PARLAYODKGVLHNEVKSILMRGLPNVYTSALSLPNIRKPDGY 1327
Db 303 RAISPQVHTIRIHYTPVRAVYODKGVFLHNEVKSILMRGLPNVYTSALSLPNIRKPDGY 362
QY 1328 DYAFSKDOYNNIDVPSRTARAITTRSGOTLSKIWNCP 1366
Db 363 DYALSKDOYNNIDVPSRTARAITTRSGOTLSKIWNCP 401
RESULT 5
Q9NA57 PRELIMINARY: PRT: 1079 AA.
AC Q9NA57;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

```

RA Borkova D., Botchan M.R., Bouck J., Brakstein L.D., Bolshakov S.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dalkle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Borkova D., Botchan M.R., Bouck J., Brakstein L.D., Bolshakov S.,
Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dalkle C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Houston D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li Z., Liang Y., Lin H.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mosheret A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munzy D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclele J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 RA Stryckas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "the genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003495; AAC22353.1;
 DR FlyBase: FBgn0030561; CG5228.
 SQ SEQUENCE 1049 AA; 107278 MW; 954D629E7619671 CRC64;

Query Match 13.0%; Score 946.5; DB 5; Length 1049;

Best Local Similarity 34.4%; Pred. No. 1,1e-53;
 Matches 376; Conservative 69; Mismatches 480; Indels 169; Gaps 56;

OY 68 EVHNPSPSSSKKAPPSGASQITKSTKSPKPNKKTKYIESEBITEHVSSENOE 127
 DB 33 EIGTAKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPT-----EGTAKPTTLKPT 86
 OY 128 SSSSSSSSSSTIMKSSKNSAANRELCKLVKDKKKRKKPPKPPVDEAGSG 187
 DB 87 GTAKPT-----LKPTGTS-----KPTTLKPT-----GTAKPTTLKPT-----EGTS 128
 OY 188 LNDGDFKVTPTDTS--TTQHNKVSSTPKITAKPINPSPSLPNSDSKETSITVNEKT 245
 DB 129 AKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTSAKPTT 185
 OY 246 VETKATTTNNKOT--STDGKETSIAKT--QSIKETSIAK--DLAPISKVLAKPT---PK 296
 DB 186 LKPTGTSAKPTTLKPTGTS---TSAKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPT 241
 OY 297 AETTKGPAALTTPKEPTPTTPKEPASTPKPEPT--PTTIKSAPTPKEPA---PTTTKS 350
 DB 242 EGTAK-----PTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTS 297
 OY 351 APTTPKEPAPTTPKEPA---PTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 407
 DB 298 AKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTSAKPTT 352
 OY 408 TTPKEPAPTTPKEPT--PTTPKEPAPTTPKEPAPTTPKEPA---PTPAKKAPPTTPKEPAP 462
 DB 353 TTKLPDGTGTAAPTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTG 412
 OY 463 TTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPA---PTTTKSAPTTPKEPSPTT 519
 DB 413 TTKA-----PTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTS 466
 OY 520 KEPAPTTPKE---PAPTTPKAPPTTPKEPAPTTPKEPAPTTPKAPAPAPKE---PAPT 573
 DB 467 AKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTSAKPTT 525
 OY 574 TPKETAPPT--PKLTPPTPEKLAPPT--PEKAPPTPEELAPPT--TPEEPPTTPKEPA 627
 DB 526 TTKPTGTAAPTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTS 585
 OY 628 PTPPKAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKETAAPTTPKGTAPPTTLKKEPAPTTP 687

DB 586 TAKPTTLKPTGTSAPPTTLK-----PTEGSAOSTTLKPTGTTAK-----PTTLK---PTEG 636
 OY 688 KKPAPKEIAPPT-----TTKEPTSTSDKPAAPTPPKGTAAPTTPKEPAPTTPKEPAPTTP 740
 DB 637 TSAOPTTLKPTGTAAPTLKPTGTSAKPTTLKPTGTSAKPTTLKPTGTTAK---PTTLKPDGTGA 691
 OY 741 KGA-----PTTLKKEPAPTTPK---KPAKELAPPTTKGPTSTSDKPAAPTPK 786
 DB 692 KPTTLKPTGTSAPPTTLKPTGTSAPPTTLKPTGTSAPPTTLKPTGTSAPPTTLKPT 751
 OY 787 EAPAPTTPKEPAPT--PKKAPPTPEPTTPSEVSTPTTK-----EPTTIKSPDE 837
 DB 752 EGTITAK-----PTTLKPTGTSAPPTTLKPTGTSAPPTTLKPTGTSAPPTTLKPT 806
 OY 838 SPPELSAEPTP--KALENSPKEGV--PTTKT---PAATPEVTTAKKDTERRDLRTPE 891
 DB 807 RT---SAOPTTLKPTGTAAPTLKPTGTSAPPTTLKPTGTSAPPTTLKPTGTTAKPTT---LKPT-E 858
 OY 892 TTTAAKMKKETATTEKTESKITATTQVSTTQDTTPKITTTLKTTTAPKVTYTK 951
 DB 859 GTSAPKPTTLKPTGTAAPTLKPTGTSAPPTTLKPTGTSAPPTTLKPTGTTAKPTT 914
 OY 952 K--TTTTIEMKPEETAKPKDRATNSKATTPKOKPTKAPK---KPT--STKKKTYMPR 1004
 DB 915 KPTTLKPTGTSAPPTTLKPTGTSAPPTTLKPTGTSAPPTTLKPTGTTAKPTT 973
 OY 1005 VKKPTTPPT---PRKMTSTMP--ELNPTSRIAEMLQTTTRPNQTPNSKLVENPSESDAG 1060
 DB 974 TEGPSAKPTTLKPTGTSAPPTTLKPTGTSAPPTTLKPTGTTAKPTT 1031
 OY 1061 GAGG-ETPHMLLRP 1073
 DB 1032 QANPFETKKERRP 1045

RESULT 7
 ID 076894 PRELIMINARY; PRT: 1795 AA.
 AC 076894;
 DT 01-NOV-1998 (Tremblere, 08, Created)
 DT 01-NOV-1999 (Tremblere, 12, Last sequence update)
 DT 01-JUN-2001 (Tremblere, 17, Last annotation update)
 DE EG:5667.1 PROTEIN.
 GN EG:5667.1 OR CG14796.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amandalides P.G., Scherter S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bekendale J., Bhandari D., Bolshakov S.,
 RA Beeson K.Y., Benos P.V., Bernasconi B., Brokstein P., Brotlier P.,
 RA Burkova D., Butchman M.R., Bouck J., Brokstein P., Brokstein P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dudin K.J., Evangelista C.C., Ferraz C., Ferraz C., Ferraz S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,

Query Match	12.2%	Score 891.5;	DB 5;	Length 1795;
Best Local Similarity	37.0%	Pred. No. 7.5e-50;		
Matches 317;	Conservative 154;	Mismatches 510;	Indels 195;	Gaps 40;

Db	624	TASITTKK----	TTSPKXITNDIDPITSTSKLSTTQKTTTTNKKTAATSTEEKPTT	679
QY	426	PKERAPITTKERAPITTKERAPATLAKKARPTTKERAPITTKERAPITTKERAP	485	
Db	660	TEKISTVSTTKTKTESSPKSTSGKPTTTPKSPSTPTPTTKKVVITVITQITTTPLRSS	729	
QY	486	PTTKSAPTTKERAPITTKSAPITTKESPSTTTTKERAPITTKERAPITTK-----	536	
Db	740	TEITSTQRPITTPQRPITTTTTLVITPKITSTTTTTEKPTTSSKRPITTTQKTTSTAMPNT	799	
QY	537	KPAITTKERAP-----	583	
Db	800	KVAITTKETPTTOSTSTTITFKTKTTNNDEPSTEEKPTSTTPKSTTTPKSTVASSV	859	
QY	584	KKLPPTPKAPITTKPKP-----	635	
Db	860	EKTISSPKPTTEKSTENPPTTNVKSALTSSTQKASTVSEPTKPT--QAITTTTKPTT	918	
QY	636	PNTPKERAPITTKERAPDT--	693	
Db	919	LKISTQBATSTQKVSIVTITTKKATBSSPLTLTSEBEPNTTKRLRTTPTTTSTVATIT	978	
QY	694	ELATTTKEPT--STSDKRAPITTKGATATTKERAPITTKERAPITTKERAPITTKER	751	
Db	979	RITTTTSESSTETTSQKPKSTPTSTPTTKPVTVTVIVSTQNPPTTTSTKSTVIT---	1033	
QY	752	APITPKKPAKELAP--TTTGKPTSTSDKRAPITTKETAPITTKERAPITTKKRAPITPE	810	
Db	1036	--TTP--NPSTQRPITTTTQPTSTASTSTISGTIRLPTTNPQNSTSDTLTVTTRPC	1099	
QY	811	TPPTTSEVSTPTT--TKE-----	850	
Db	1093	PDPTSDKNTNACTQALDQVNLLELQSPQKQEOPTNTRTALGSRNLTGGQGVPRY	1155	
QY	851	LENDPKRPGVPTTKTRATPEKMTTAKOKTTER---	907	
Db	1153	MDDPSSAEESGATATKAPATSTLAAHLLQKLRHLITSTTPRSREHAP--TORPSSOP	1210	
QY	908	EKTTESKITATTTQVSTTTQDTPPKITTKTT-----	963	
Db	1211	SSSQRSR--GVTIQAMRHNLTATSKPFIANSRLRSTIQQLASTQKRSITPTVITNHTKPE	1266	
QY	964	ETPAKPKDRATNSKATITPKQAKPTKA--PKKPTSTKKPKTMPRVKPKTPTTPPKMTSTMP	1022	
Db	1270	ED--SEYDSETSQYDDEDEVDLKDQPRAMSSITVAALPAPVSTPTTEREPQK--TSSSP	1327	
QY	1023	ELNPTSKRIEAMLOTTTRPQGNQNSKLVEVNPSED	1058	
Db	1328	--SPT----KATSSITQPTETTTGDLLEVDSSGSSD	1357	
RESULT	8			
ID	096449	PRELIMINARY:	PRT:	1489 AA.
AC	096449:			
DT	01-MAY-1999	(Tremblrel. 10, Created)		
DT	01-MAY-1999	(Tremblrel. 10, Last sequence update)		
DT	01-MAY-1999	(Tremblrel. 10, Last annotation update)		
DE	CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.			
GN	CAR99.			
OC	Phytophthora infestans (Potato late blight fungus).			
OC	Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;			
OC	Phytophthora.			
OX	NCBI_TaxId=4787;			
OX	NCBI_TaxId=4787;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=RACE 1-11;			
RA	Goerhardt B.;			
RL	Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.			
SR	EMBL: AF061185; AAC72308.1; --			
SO	SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;			

Query Match 11.8%; Score 863.5; DB 10; Length 1489;
 Best Local Similarity 31.7%; Pred. No. 4.1e-48;
 Matches 368; Conservative 55; Mismatches 509; Indels 229; Gaps 51;

107 TKKVIIESEETIEHSVSENOSSSSSSSSSTWIKKSSKNSAANRELÖKKILKYDNK 166
 282 TPVYGYSTEETEGOHVGTGEPSEDETAPTEGTTY--VPREETIAASE-----DTIYAP 334
 167 KNRTKKRPTRKPPVVDAGSGLDNGDKVYTPDTSTT-----QHNKVSSTPKITTKAP 220
 335 REVTPVAPTEKPYDVEETIYVTEETIYAPTKSETNAPTEHMYAHIEKPCDTEVTIYAP 394
 221 NPKRSLP-----PMSDT-----SKETSILVKEETVEKEKTT--TNKQT 258
 395 EETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTE 454
 259 STDGEKTSKAKETOSIEKTSADLAPSKVLAKPTKAEETTKGPAITTPKEPTTPPK 318
 455 YAPTEETIYAPTEETIYAPTEETIYAPTKETIYAP--EETIYASTEETIYAPTEETIY 511
 319 EPASTTPKEPTPTTKSAPTPPKAPAP--TKSAPTPPKAPAP--TTKEPAPTPKEPA 375
 512 APAEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPAEETIY 570
 376 PTTKKEPAPTTKSAP--TPKEPAPTPPKAPAPTPPKAPAPTPKEPTTP 425
 571 EETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPAEETIYAPTEETIYAPTE 629
 426 PKEP-----AP--TTKEPAPTPKEP-----APAPKAPAPTPPKAPAP 462
 630 TYAPTEETIYAPTEETIYAPTEETIYAPAEETIYAPTEETIYAPTEETIYAPTEETIY 689
 463 TTPKEPAP--TTKEPSTTPKEP-----APT--TKSAP--TTKEP-----TTKEP 499
 690 TEETIYAPTEETIYAPAEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTE 749
 500 --APT--TKSAP--TPKEPSTTPKEP-----APTTPKEPAPTPPKAPAPTPPKAPAP 548
 750 TYAPTEATYAPTEETIYAPAEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIY 809
 549 T--TPKEPAPTPPKAPAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTP 606
 810 TEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTE 867
 607 PEELAP--TPKEPSTTPKEP-----APT-----TPKAAAPNTKEPAPTPPK 648
 868 ETTAPTEETIYAPTEETIYAPTKETIYAPTEETIYAPTEETIYAPTEETIYAPAEETIY 927
 649 EP-----APTTPKEPAPTPPKAPAPTPPKAPAP--TTKEPAPTPPKAPAP--ELAPT 698
 928 EPTTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPAEETIYAPTEETIY 987
 699 --TTKEPSTTSKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAPTPPKAPAP 746
 988 ETTAPTEETIYAPTEETIYAPTEETIYAPAEETIYAPTEETIYAPTEETIYAPTEETIY 1047
 747 -----TTKEPAPTPPKAPAP-----APKE--LAPT--TTKGPSTIYSDKAP 782
 1048 ASTEETIYAPTEETIYAPAEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAP 1107
 783 TTPKETAPTTPKEPAPTPPKAPAPTPPETPPPTSEVSTPTTKAPT--TIHKSDESTP 840
 1108 ETTAPAEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIY 1166
 841 ELSAPETPKA--LENSPKREG-----VPTTKPAAATKPKPETTAKKKTTERDLRTPP 890
 1167 YAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIY 1226
 891 ETTAPAKMTKETATTEKTESKITATTOYSTTODTT-----PKKIT-- 937
 1227 EETIYAP-----TEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIY 1281
 938 IKTTLAPKVTTKTITTEIMNKPEETAKPKDRAT--NSKATTPKP--QKPTKAPKKPT 994

DB 1282 TEETIYAPTEATIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAP 1341
 QY 995 -----STKPP-KTMPVVRKPKTTPPRKMTSTMPLEINPISRAEAMLOTTTP 1041
 DB 1342 PYPAEESTSTVSTERPCNTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIY 1399
 QY 1042 --NOTNPKIYEVNPKSEADG 1060
 DB 1400 CDMOGINGICVENKVRINNG 1420

RESULT 9
 ID Q20007 PRELIMINARY; PRT: 1274 AA.
 AC Q20007;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE COSMID F35A5.
 GN F35A5.1
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favellio A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rinken L., Roopra A., Saunders D., Showmken R.,
 RA Smalton R., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RC Leimbach D.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U46675; AAB52641.1;
 SQ SEQUENCE 1274 AA; 138065 MW; 07F6BD292C1799F CRC64;

Query Match 11.6%; Score 846; DB 5; Length 1274;
 Best Local Similarity 27.6%; Pred. No. 4.8e-47;
 Matches 313; Conservative 119; Mismatches 455; Indels 248; Gaps 60;

QY 76 PSSKAPPPSGASQITKTRSP-----KPP-----NKKTKKVIIESEETIEHSVSENOES 128
 DB 185 PSPKKAAPSKHDPVPTPIKKNPAKKKPPWEDDEVTEIKKPEPATRKVPALKKKEP 244
 QY 129 SSSSSSSSSSTWIKKSSKNSAANRELÖKKILKYDNK-KNRTK-KP-----TP----- 176
 DB 245 STSVKPVSPDPPTKKV-----PVKKEPEVPPPIKKNPAKKKPPWEDDEVTEIVK 294

```

QY 177 KPVVDEAGSLDNC-----FKVTPDTSTQHNKVSTSPITTTAKPI-NRPSLPPNSD 231
Db 295 EPPVPEKAPVLLKKDPAPAKARDPSPSKAPPKVESSPVVPTPTKANKPKKPPWE 354
QY 232 TSKETSLTVNKETTVEETTTNTKOTSTDKKTKTSAKETQSIEKTSKADLAPTSKVL 291
Db 355 VDDEPAEVEVKPSAPEKTPVLKKKEPPSSTSSDPSKKAAPAVKPRDSSPKKATPL 414
QY 292 KPTPKAE-----TTTKGPA-----LTTPKKEPTPTT-----KEBASPTPK 326
Db 415 QADBEKAOEVPPTPVKNPVKKYKPPMEVDEDEPVEEVKQPEAPAKTKPLKKKEPAADTA 474
QY 327 EP-----TP-TTKSAPTTPKKEPAPTTTKSAPTTKEPAPTTTKPA----- 367
Db 475 KPAVSKPPTPEKKDPKPRDSSPKKVAAKDSQAQPA-TVKNPNVKKWPRPMEDEDETPA 533
QY 368 -----PTTPKEPAPTTTKEPAPT-----TTKSAPTTPKRP-----APTTPKKPAPTTP 410
Db 534 DQVSKPPTDAKTPSLAKDPAKAPAKESLKPADTKAPAKPRDSSPKKVAAPAKPKTPVLA 593
QY 411 KE-----PAPTTPKKEPTPTTPKEPAPTTKEPAPTTPK-EPAPTTAPKKPAP-----TTPKE 459
Db 594 KKEBAGPADSKTKEPEKSKPRDPSKKAAPKAVPKTEVNAAPAKKKEPISKPKDTAPK 653
QY 460 PAPTTPKEPAPTTTKEP-----SPTTPKEPAPTTTKSAPTTTKEPAPTTTKSAPTTTP-KEP 514
Db 654 AEPNSPVVP-PTPVKNPNVKKWPRPMEDDDAFAKPVSLPEPEKK-TPLAKKAKPPDSEA 711
QY 515 SPTTKEPAPTTTPK-EPAPTTPKKPAP--TTPKEPAPTTTKEPAPTTTTPKAPTA-PKE 569
Db 712 AADVSGPSSADPKLAKAPKAPKPRDPSKKAAPKAPTKAPAKT--EVPVAVVKKPEVAAKSRD 769
QY 570 PAPTTPKEPAPTTPKKLLPTTP-----EKLAPTTPEKAPTTPEELAPTTPEEP-----T 619
Db 770 PSPKKAK-AEENSB--VVPPTPVKNPNVKKWPRPMEDDDAFAEPVNPVEPEKKTPVLAKKT 826
QY 620 PTTPKEEAPTTTPKAAAPNTKEPAPTTTPKEPAPTT-KEPAPTTTKEPAPTTTPKEGTAAPT 678
Db 827 PVKRDSPKKAAPKAPKSTTDADPVVKKPEVSKPEKSPSKKAENSPVVP---PTP 882
QY 679 LKEPAPTTPKKPAKELAPTT-TPKEPTSTSDKAPPTTPKGTAPTTKEPAPTTTPKEPAP 737
Db 883 VKNVKKW-KPMEDEDEPEVEVKPSE--PEKTPVLAK-KEPEKKD-APKAAKAPR 937
QY 738 TTPKGTAPTTLKEPAPT-----TPKKPAP-----KELAPTTTGGPTS-----TTSK 779
Db 938 PSPKKAPE--KEPAKVAAPRDLSPKKAIPANTQEAPPTPVKNPNVKKWPRPMEDEDE 995
QY 780 P-----APTTPKEN-----APTTPKEPAPTTTPKKAAPTTPPETTPPTTSVSTPTTTPKEP 828
Db 996 PAEVSAPPEPEKTPVLAKKAPAKPRD---SPKKAAPVAAK-PDPKIPV-PPTPVKNP 1050
QY 829 TTIHKS-----DESTPELSA-EP--TPKALENSPKKEGV-----PTTKTPA 867
Db 1051 VKKMPMEDEDESEVSAPEPEKTPVLAKKAPTKAPKAPKPOSEAAADVSGTSPKSDPK 1110
QY 868 AT-----KEEMTTAAKDKTTERDLKTTPETTT--TAAPKMTKETATTEKTSKITATTT 919
Db 1111 LSKKAPVEKPKPTTDPKDDDLKSPAPKAPKAPAKPKKVPVDDDDPEADDFVPA 1170
QY 920 TQVSTTQOTPEPKITLTKTTLAPKVTTTKTITTTTELKMPBEIAKAKDKDASATNSKAT 979
Db 1171 PSKKPTEDEADPLG-----GPKTKDPK-----LNKKAPEKTEK----- 1206
QY 980 TPKPQKTPAKPKPTSTKPKKPTMPKAPK-----TTPPKKMTSTMP 1022
Db 1207 -PKPEVSKPEPKPTPEPKP-AAPKPKWPRPMEDDDPEADFTMPAPKPKDTEDP 1259

```

```

RESULT 10
Q9SPMO ID Q9SPMO PRELIMINARY; PRT; 1315 AA.
AC Q9SPMO;

```

```

DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 17, Last annotation update)
DE EXTENSIN-LIKE PROTEIN.
GN pex2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OC NCBI_TaxID:4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=POLLEN;
RA Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA Hohorst D., Gao M., Showalter A., Bedinger P.A.;
RT "Pollen Extensin-like (Pex) Genes in a Monocot and a Dicot.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159297; AAD55980.1; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR001998; Xylose_isom.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR003592; LRR_out.
DR Pfam; PF00360; LRR; 3.
DR PRINTS; PR01217; PRICHEXTENS.
DR PROSITE; PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR SMART; SM00370; LRR; 3.
SQ SQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F C664;

```

```

Query Match 11.2%; Score 817; DB 10; Length 1315;
Best local Similarity 23.4%; Pred. No. 3,9e-45;
Matches 269; Conservative 109; Mismatches 490; Indels 282; Gaps 37;

```

```

QY 22 VSSQDLSSCAG-----KGEGTSKDATCNCIDYCNQHYMECCDFEFCVCTAAEVHN 71
Db 369 VNEDDRNCAGALRPAKTAALQCAPVLARVEVDCS-----KHVC--AGYPT 413
QY 72 PTSPSSK-KAPPSGASQTIKSTKSPKPKPKKTKVIESEETIEHSVSENOSSS 130
Db 414 PGGPSSSVGPKRPSVVGK-----PAAPAPMPTPHPPVSPPELPPEPVP----- 459
QY 131 SSSSSSSSTIKIKSSKSAANRELOKLLKVDNKNKTKKPKTPKPVVDEAGSLDN 190
Db 460 -----PAPAPMPTPLRSPADEYI-----PYPVAPKSPG--- 491
QY 191 GDFVTPDTSTQHNKVSTSPKITTAAPINRPSLPPNSDTSKETSITVNKETIYETKE 250
Db 492 -----TSPASRGA-PPLAQAPPAASSPAPTPVKSPPPAV----- 527
QY 251 TTTTNKOTSTGKEKTTSAKETOSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTPK 310
Db 528 -----VLPRAKTTSPAPV-ASPAPAPVSSPPQVQKSP 562
QY 311 EPPPTTPKEPASTTPKEPPTTITKSAPTTKEPAPTTTTSAPTTTP-KEPAP-----TT 362
Db 563 PPAPVASSPPPMKSP--PPAPVASSPPLKSSPPAPVASSPQPLKSSPPVLMSTPS 620
QY 363 TKEPAPTTKEPAPTTTKEPAPTTTKSAPTTPE-----PAP--TTP-----KKAPTTPE 412
Db 621 VKSPPPPVAVASPPPVKSPPLAVSSSPVKKLPLPLPAGSTPPEEKEKTPPTPVK 680
QY 413 PADTPKE-PTTPTPKEPAPTTKEPAPTTTPKEPAPTPAPKAPATTPTPKEPAPTTTPKAPAT 471
Db 681 SSPPPKSLPPPLTSSPQOEKPTPSPISKPPSPVETLPPSKSSPPEEPVSSPPQ 740
QY 472 TTPKEPPTTPKEPAPTTTTSAPTTTKEPAPTTTKSAPTTTPKEPSPPTTTPKEPAPTTTPKEP 531
Db 741 APKSSSP-----PAPVS--SPPLKSSPPVPPESSPPPTPKSSPPLAVSSSPQVYEKTS 793
QY 532 PTPPKKAPATTPKEPAPTTTPKEPAPTTTTPKAPTAKEPAPTTTPKETAATTPPKKLLPTTP 591
Db 794 PPAPVSSPPTPKSSPPLAVSSSPQVEKTSPPAPVSSPPTPKSSPPLAVSSSPQVE 853

```

```

QY 592 EKLAPTPEKAPPTPEELAPTTPEEPTPTTPEAPPTKAAAPNTKPEAPTTKEPA 651
DB 854 KTSPPAPVSSPPLEPKSSPSVSSPPTTVSSPPAPLSSPPAKSSPPAHVSSP 913
QY 652 PTTPEKAPPTPKETATTTKGTGTA-----PTTKEAPTTTTPKKAPKELAPTTT 700
DB 914 PEAKSSPPLAIISSPSEKSPSSPPMVKETSPPRATVSSPPTTPKSSPP---AVSS 970
QY 701 KEPTSTSDKAPPT---TPKGTAPTT---KEPAPTTPKPEAPTTTK---GTAP 745
DB 971 PPAVKKSSPPAPVSSPPPTPKPLPPAPVSSPPVKKSSPPPTPVSSPPTPKPLPPP 1030
QY 746 TTTKEAPPTPKKAPKELAPTTTKGPTSTSDKAPPT---PKETAPTTTK---EPAPT 799
DB 1031 TPVSSPPTKPLPPP---AVSSPPVVKSSPPAPVSLPPTPKKSSPPTKRVSSRPV 1087
QY 800 TPKKAPPTPEPTTSEVSTPTTKEPTTHKSPDESPTPELSAEPPTKALENSPKEPG 859
DB 1088 VKCSPPTIVSSPPAPKSLPPPTPVSSP-----PREVKSPPPTPVSSPPAPK 1137
QY 860 VPTTKTPAATKPEMTTAKDKTTERDLATTP---ETTAPKMTKETATTTKTESK 914
DB 1138 SSPPPTPVSSPP-----ELKSSPPAPVSSPPAPKSSPPAPVSLPPEVK 1185
QY 915 ITATTTQVSTTTQDTTPEKITTLTKTTTAPKVTTKTITTTIIMKKPEETAKPKDRAT 974
DB 1186 SSPPAPLSSPPPAKSP-----PPPA-----PKMSLPPPKSP 1219
QY 975 NSKATTPKPKPTAKPKKPKSTKPKTWPVRKPKTPTTPKKMTSTMBELNPTSRILAEAM 1034
DB 1220 PPAVSSPPPKSPSPAPLSSPPAP---VKPPSLPPAPVSSPPAPV----- 1267
QY 1035 LQTTTRPQGTNSKLIVENPSEDAAGEGEPTPHMLLRPHVEMPEVMDMYLPRVNOG 1094
DB 1268 ---TSAP-----PKEDSTRA-----PPAALPPPSIND 1293
QY 1095 IINPMLSE 1104
DB 1294 IILPPIMANK 1303

RESULT 11
Q9LIE8 PRELIMINARY; PRT; 1480 AA.
ID Q9LIE8
AC Q9LIE8:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI TaxID=3702;
OX [1]
RN RN SEQUENCE FROM N.A.
RP RP STRAIN-COLUMBIA;
RC RC Kameko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RX RX Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RA RA [2]
RA RA SEQUENCE FROM N.A.
RA RA STRAIN-COLUMBIA;
RX RX PubMed-10907853;
RA RA Nakamura Y.;
RT RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT RT Sequence features of the regions of 4,251,695 bp covered by ninety PL,
RT RT TAC and BAC clones."
RL RL DNA Res. 7:217-221(2000).
DR DR EMBL: AP001306; BAB03062.1;
DR DR InterPro: IPR002965; P. rich-lexensn.
DR DR PRINTS: PR01217; PRICHEXTENS.
SQ SEQUENCE 1480 AA; 147153 MW; DIACOC79F155E732 CRC64;

```

```

Query Match 11.2%; Score 814; DB 10; Length 1480;
Best Local Similarity 28.8%; Pred. No. 6.9e-45;
Matches 329; Conservative 61; Mismatches 541; Indels 212; Gaps 50;

QY 66 AAEVHNPTSPSSKAPR---PSGASQITKSTTKRSKPRKPKTKTKVLESEITEHSV 122
DB 51 AVKPHNPKRPETIKRPPRKKNHPRKPTVK-----PHNPKRPRTKNHPRKPTVK 102
QY 123 SENOESSSSSSSSSTIKKISSKNSANRELQKKLYK---DNKNKTKKKTKPKRV 180
DB 103 -----PHKRPETIKRPHNPK---PHNPKRVNKNRPHNPKRTKRNHKNR 146
QY 181 VDEAGSLGDGDKVLTDRSTTOHNKSVSTSPKITTAKRINPRASLPNSDTSKETSLAY 240
DB 147 -----TKHN-----PHNPKRTIKRPPKRVSVKPRPSTPK----- 175
QY 241 NKEETVEKKEPTTNKSTODGKEKTSAKETOSIEKTSKDLAPISKV---LAKRPKAE 298
DB 176 -----PTTNPPSPTRPRTNKRPRCTRPVASSPMATPPTOMRPIATP-PIAK 224
QY 299 TTTKGPALTTPKEPTP---TTP---KEPASTPKKEPTPTTKISAPTTPEKAPPTTT---K 349
DB 225 SPVATPPIATPPTATPPTITPVPATPPTTPPIANPPIINPPIATPVPVAPPTINPISK 284
QY 350 SAPTTPEKAPPTTT-----KEPAPTPEKAPPTTKEBA-----PTTK 388
DB 285 PVTATP---PTTTPPIAKRPPIATPPISTPPAATPPAATPPTITLPPAKRVAISPIVT- 339
QY 389 SAPTTP---KEPAPTTPKAPR---TTPKPEAPTTPEPTPTPKPEAPTTKAPPTTP---KEP 444
DB 340 -PVPATPPIAQRPVATPPIATPVPVATPPIATPPTSPISITPISBPVATPPTATSPKTP 398
QY 445 APAPKAPAPPTPEKAPPTPEKAPPTT---KEPSPTPKPEAPTTKSAPTTPKEPA 500
DB 399 PPAKPPVATPPIAKSPATPPTATPVPATPPIEKPVPATPPTTPPAT---PVAAPRV 455
QY 501 PTTTKSAPTTPKPEP---SPTTKEPAPTTPKEPAD---TTPKAPAPTPEKAPPTTPKEP- 554
DB 456 ETPPIATPPTKAPRISNPKSPKRVATPPIATPPTITPTTPVKKRVANPPIAIPVAPRV 515
QY 555 -APTTTKKAPATP---KEPAPTTPKEPATTT-----PKKLTTP---TTPKELAPTTPEKPA 603
DB 516 VTPPTAPPPATPPIAASVATPPTATPVPATPPIAKRVATPPTTPPIATPVPAPRV 575
QY 604 PTTPEELAPPT-PEEPTPTPEEAPPTTPKAAAPNT---PREAPTTPEKAPPTPEKPA 659
DB 576 ATPPIATPPIAKRPISTPPIKSPVATPPIAPATPPIITTPPAKRPVATPPIATPPIAKRV 635
QY 660 PTTPEKETAP---TTPKGTAPTT---LKEPAPTTPKAPKELAPTTKEPTSTSDKAPPT 714
DB 636 ATPPIATPPIATSVAKRPVATPPIKTPPAKRPVATPPIATPVPAPRVAT---PPTAT 692
QY 715 TPKGTAP---TTPKEAPPTPEKAPPTTPKGTAPTTKELAPTTPEKAPKELAPTTTGPT 773
DB 693 PPIATPPIATPVPVTPPPAPVATPPIAKPPTTI---PPIATPVPVAPPIATPPIATPPI 750
QY 774 ST---TSDKAPATTPKETAPT---TPKEAPATTPPKKAPPTTP-----ETPPTTSEVSTPT 824
DB 751 ATPPIATPVPAPKPPVVPPTATPPIATPPIAKSPVATPPIATPVPAPPIAKPPVAPPT 810
QY 825 TKEPTTH---KSPDESPTPELSAEPPTKALENSPKEPGVPTTKTAATKPEPT---TTAKDT 881
DB 811 TAPPTAPPPVAKPPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIAPRV 870
QY 882 TERDARTPETTTAAPKMTKETATTEKTSKITATTTQVSTTTQDTTPEKITTLTKTT 941
DB 871 ATPPIATPP---IAKPPVATPPIATPPIATSPVAPVAIPKIPKPPAKPPVATPPIATP 928
QY 942 TLAEPVTTTKKT---ITTTIIMKKPEETAKPKDRATNSKATTPPKPKTAPKPKSTK 997
DB 929 PVAKPPVATPPIATPPIATSPVATPVPVT---PPTATSPVATPPIAKPPTTTP---PPIAT- 982

```


DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 3.
 DR SMART: SM00370; LRR; 4.
 DR Kunitz signal.
 FT SIGNAL.
 SQ SEQUENCE 1 27 POTENTIAL.
 1188 AA; 120980 MW; 2C77C7F8D7130149 CRC64;

Query Match 10.7%; Score 784; DB 10; Length 1188;
 Best Local Similarity 29.9%; Pred. No. 4,9e-43;
 Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

OY 174 PTKKPPVVDAGSGDNGDFKVTTPDTSTQHNKYSTSPKITTAKPDPNPSL-----PPN 229
 DB 460 PTHSPAD-----DVPPTPVPGKSPATSPQVQPPASTPPSLVKLSPPQ 510
 OY 230 SDTSKETSILVNNKETTETTTNKQSTDKGKRTSKKEOSIENKSAKLAPTSKY 269
 DB 511 APVG-----SPPPVKTTSPADIG-----SPSEPPVSVV 541
 OY 290 -----LAKPPKAPTTTGGPALTTPKEPTTPKEPASTTPKEPTTTIKSAPTTKEPA 344
 DB 542 SPPPPVSPPPAPVSGSPPEKSPPPAPVASSPPPVKSP--PPPTLVASPPPVKSP 599
 OY 345 PTTTKSAPTTP--KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 401
 DB 600 PPAPVASSPPPVKSPPPPTPVASPPPVASSPPPMKSPPPPTPVSSPPPEKSPPPP 659
 OY 402 PKKAPTTTKEPAPTT-----PKE-----PPPTTKEPAPTTTKEPAPTTTKEPAPTTAKKP 452
 DB 660 PAKSTPPEEPTPTPVSKSPPEKSLPPPTLLIPSPPOEKPTPTPTSPK--PSSSEKP 718
 OY 453 APPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 512
 DB 719 SP--PKEVSSPPQTP-----KSSPPAPVSSPPPTPVSSPPALAPVSSPSVSKSP----- 768
 OY 513 EPSPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 572
 DB 769 PPAPLSSPPAPVOKVSSPPPOVSSP--PRAKSSPPPLAPVSSPPQVEKTSPPAPLSSPP 827
 OY 573 TTPKRTAP-----TTP-----KKLTPTTPEKLAPTTPEKAP-----TTPPELAPTTPEPT 619
 DB 828 LAPKSSPPHVAVSSPPPVKSSPPAPVSSPPPLMKPAPSAHVAVSSPEVVKPSTP--PA 885
 OY 620 PTTPEEAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 678
 DB 886 PTTVISP--PSEPKSSPPPTPVSLPPPIVSSPPAMVSSP--PMTKSSPPPVVAVSSPPPT 943
 OY 679 LKE-----PA-----PTTPPK--PARKELAPTTTKE--PTSTSDKPAPTTTPKGTAPPTP 724
 DB 944 VVSSPPAPVSSPPATPKSSPPAPVNLPPPEVKKSSPPPTPVSSPPA--PKSSPPAP 1000
 OY 725 -KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 783
 DB 1001 MSSPPPEVKSPPAPVSSPPPVKSSPPAPVSSP--PPVKSPPPPAPVSSPPPV 1057
 OY 784 -TPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 842
 DB 1058 KSSPPAPVSSPPPVKSSPPAPVSS--SPPPVKSSPPPPAPVSSPPPIKSSPPAPV 1115
 OY 843 SAPPRTKALENSKKEGVTTKTPATKPEMTTAKDKTTERDLTPTPT 892
 DB 1116 SPPAP-----VKPPSLP--PAVAVSSPPVVTAPAPKKESDLPAPAS 1158

RESULT 14
 OYVEL9 PRELIMINARY; PRT; 2112 AA.
 AC OYVEL9:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG4090 PROTEIN.

GN CG4090.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RM SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Bradton R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Bonch M.R., Bouck J., Brokstein D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Boulter H., Cadieu E., Center A., Chandra I.,
 RA Burris K.C., Busam D.A., Butler H., Cadenhead L.B., Davies P.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwan C.,
 RA Jaislid M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paolel J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AEO03716; AAF55402.1;
 DR Flybase: FBgn0038492; CG4090.
 DR InterPro: IPR002557; Chitin_binding.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF01607; Chitin_bind_2; 10.
 DR SMART: SM00494; ChIBD2; 11.
 DR PROSITE: PS00022; BGF.1; 1.
 SQ SEQUENCE 2112 AA; 219547 MW; B91018E551A5D36 CRC64;

Query Match 10.7%; Score 781; DB 5; Length 2112;
 Best Local Similarity 27.3%; Pred. No. 1,4e-42;
 Matches 309; Conservative 98; Mismatches 367; Indels 356; Gaps 48;
 OY 49 NCQHYMECCPPEFKVCAAEVHNPTSPSSKKAPPGAGAGQTIKSTKRSPKPPNKKTK 108
 DB 987 NCSKFRVCDNGKGGFTKVSF--TTPNTLIMPEANSCHNPDIQK-----PLCKK 1036
 OY 109 KVI-----ESEITEHVSVENOESSSSSSSSSTTWIKSSKNSAANRELKAKLVKD 164
 DB 1037 KVSQGGSSNSTGSS 1076
 OY 165 NKNKFTKKRPKPPVVDAGSGDNGDFKVTTPDTSTQHNKYSTSPKITTAKPDPNPS 224
 DB 1077 -----GSSNSTGSSSSNSGASGSSGSSGSSSSSSSSSSSSSSSSSSSSSSSS 1103


```

OY 225 SLPPNSDTSKETSLVYKNETVETKETTNNKOTSTDKKETSAKETOSIEKTSK-CL 283
DB 1104 -----NSGSSGSGNSGNGOTS-----SSSSSSSSSSNNNNQSSSSSSSSSSSTSSKPNP 1154
OY 284 APTSKYLAPTPKRAETTK-----GP----- 304
DB 1155 SETCKVNGOIFIDRSDCAFYRCVNDRCGFNNVFPSCGPTGVMDAQMOCNHNAAVKEC 1214
OY 305 -ALTTPKEPTTPPKPEASTTPKEPTTPKSAPTTPPKPAETTKSAPTTKKEAPTTT 363
DB 1215 GGIAPPTSTPTTSR-PTTASTSRSDOTSTSRPGP---PTAKPVYARPTTSSPTTA 1269
OY 364 KEAPPTKEAPPTT-----TKEP-----APT----- 385
DB 1270 SSSQTTSPVTOAPNTDGKCRSEGFMAADPNNSCFYRCVANNKGFSTSLPFOGACTVMDQ 1329
OY 386 -----TTKSAPTKEAPPTTKKAPPTTKPEAPTTTKEPTTPKPEAPPTT 433
DB 1330 DLQTCNNHNNNCSTGESTETPKRPC--EPATNGTATSTSTTPP--PPTTDLPTSTYT 1385
OY 434 KEAPPTKEAPTAAPKPAETTKPEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTT 493
DB 1386 GLP-PTTTTELPT-----TTDLPTTTRLRPTTTSLPPTTTGLPPTTTGAOP 1437
OY 494 TTTKEAPPTTSAPTTK---ESEPPTTKPEAP----- 524
DB 1438 TTTTSSSETSTVTSTSESTTQPSSTTMMKPLPAGTECTEGVMADEDCRKYRCINA 1497
OY 525 -----TPPK-----EPAP-----TPPKKAPPTP-----KEAPPTTPE- 553
DB 1498 GASRYKYNFTCKGNGNEEYOTCYVNIIPRCSKLPAETPTTTESESKBGSTPOST 1557
OY 554 PAPPTTKKPAETKPEAPPTTKETAPPTPKLPPTPEKLAPTTPEKAPPTPEELAPT 613
DB 1558 DEPTVTKPI-TKPT-EPSTEKPKPTQYBEKPTTEE--PEKPKPT-TTEYPOKPT 1612
OY 614 TPEETPTTPEAPPTTKAAPTTPKPEAPTTTKEAPPTT-PEKAPPTTKETA----- 667
DB 1613 TTEETPTPKPKPPT-TTEYPOKPTTTEETPEKPKPTTTEYPOKPTTTEETTSIDGYN 1671
OY 668 PTPPK-CTAPPTTKEAPPTTKKAPKELAPTTTKEPTSTSDKPAETTKGAPPTPK 725
DB 1672 PTTTSVPCYNPTTPIPETTSTPGK---PTTGTBITTT-LPSTTTDAIOEPTTSK 1727
OY 726 EPAPPTTPE-PAPPTPKGATPTTKPEAPTTPK-----PAKELA-----PTTKGPT 773
DB 1728 KPEPTTTESESTPEGCVTTLOPEOPNYNCSSEGEFFPDPCSRHYRCVDAKNGKY 1787
OY 774 STSDKPAETPKETAPPTTKEAPTTKPAETPTTPEPTTSEVSTPTTKETPTTHK 833
DB 1788 QVAFKCGKGVMDTSTETCVADQVGN-----CSSQOTTPTTTEPTGTTES 1836
OY 834 SPDESTPELSAETPKALENSPKGVPPTTKPAATKPEMTTAKDKTTRDLATTPET 893
DB 1837 TTSSGKP-----ETTSKAPEN-----TTTNA--PETTTT-----SESETT 1869
OY 894 TAAPKMTKETATTEKESKITATTTQVSTTQOTTPKITTLLKTTTAPVTTTKT 953
DB 1870 TT---VASETTTTSGT-----TTTANPETTTKPKP-----ETTITAGEETSTSKS 1913
OY 954 ITTTEINNKBEETAKPRDATNSKATTPKPKPTKAPKPKTSTKKPKTMP 1003
DB 1914 PTTTE-----SPAPTSTMSAP-----CEETCP 1935

```

```

RESULT 15
O9XDH2 PRELIMINARY; PRT; 763 AA.
AC O9XDH2;
DT 01-NOV-1999 (T-Emblrel. 12, Created)
DT 01-NOV-1999 (T-Emblrel. 12, Last sequence update)
DT 01-JUN-2001 (T-Emblrel. 17, Last annotation update)

```

```

DE PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Espilite C., Lacleite J.L., Mondragon-Palomino M., Martens A.,
RA Zhang Y., Moreno C., Singh M.;
RT "Cloning and characterization of a new member of the PGCS family that
RT is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF071081, A041594.1; -
DR InterPro: IPR002951; Atrophin.
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PRO1227; ATROPHIN.
DR PRINTS: PRO1217; PRICHEXTENSIN.
DR PRINTS: PRO1218; PSTLEXTENSIN.
SQ SEQUENCE 763 AA; 75034 MW; 39168EC45A5916F8 CRC64;

```

```

Query Match          9.8%; Score 715; DB 2; Length 763;
Best Local Similarity 28.2%; Pred. NO. 9.6e-39;
Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

```

```

OY 293 PTPKAEETTKGAPALTTPKEPTTPPKPEASTTPKEPTTPKSAPTTKPEAPTTTKSAP 352
DB 3 PVP-----APRALDLPAPAPAPAPKSKPPPPAP-----PAPCKMLVSAAP 46
OY 353 TTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTSAPTTKPEAPTTTKEAPTTTKEAPTTT 412
DB 47 PCP-PAP-----PAPKPKSKAPFPVPPAPAPKELAPLP--PAP-----PEAPRE 90
OY 413 PAPTTKEPTPTT-----PKPEAPTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAP 466
DB 91 SRALPCCPBPVVIDPPEAPAPVPPAPNSPPFPFPAPKFFVAPVVP--PVNSPP 148
OY 467 EP---APTTESEPPTTKEAPTTTSAPTTKPEAPTTTKEAPTTTKEAPTTTKEAP 524
DB 149 FPPFPALNPPAP-----PAPPLANSPLPAPAPPPAGT--PPAPWPVPVAPAPSKXA 201
OY 525 TTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAP 580
DB 202 SPKRPAP-----PMPATKEFPPLPVPDPISKETPPAPAPPIPPAPVPIPPVPLP 256
OY 581 TTPKLTPTTPEKLAB-----TTPKEAPTTPEELAPTTPEEPTPTTPEEAPPTT-- 630
DB 257 PVPNKKIPAPP--APPVAVAVLAPCPPLPLPNNHPPAPAPAPVGPVLAFLPNSHP 313
OY 631 ---PKAANTPKPEAP----- 644
DB 314 PAPPASVPGVLAFLPISGRVSVWKGSTTLSTFCCHVCSGEVLALANDSRSRSLP 373
OY 645 --TTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTT 694
DB 374 TTTTALPAPIPPLPPLPPLPINTAVPPIPLPPTTALAAPLPPLAPLPLISGVAPAP-- 431
OY 695 LAPTTTKEPTSTSDKPAETTKGAPTTPKKA-----PTTKEAPATTPKGAETPTTL 748
DB 432 --PIPPGKMTTPIPLAPAPBEKP-TVVVLPPGSPCSPSEKPNPAPPEPEPEKSSPALP 488
OY 749 KEAPPTP---KKPAPELAPTTTGP--TSTSDKAPATTKRET---APTTPKEAPTTT 801
DB 489 APPAPSMBSAVVPPSPPIPPAPAPAPASMPALPAPPSPAPATLCPPLPSPAPASP 548
OY 802 KKPAPTTPEPTTSEVSTPTTKETTHKSP-----DESTPELSAETPKA--LENSP 855
DB 549 --PAPAPPTTPEKLS--ANPECCPVPVAPNPPAPAPAPAPAPAPAPAPAPAPAPAP 604
OY 856 KPPGVPPTTKPAATKPEMTTAKDKTTRDLATTPETTTAAPKMTKETATTTETTESKI 915

```

```

Db      605  -PAPAPAPPPSALPEVNPDA-----PPTPAAK-----SRAL 637
Qy      916  TATTQVSTTDDTPEFKITLTKTTLPAYTKITITTEINNKDEETAKPKDRATN 975
      :      :      :      :      :      :      :      :      :
Db      638  PAAPAPAPPVATRTT-----PAPAPAPAPN 665
Qy      976  SKATTPEKPKP-----TPAKPKPTSTKKPKTMPEVRKKTPTPKMTSM-----PELNP 1026
      :      :      :      :      :      :      :      :      :
Db      666  SMAPEAPAPDPDPLLATPAPAPAPPLPMSPPAPPLPAPAPPPAPPLTINOPSPPLAP 725
Qy      1027 TSRIAEAMLTQTTTRPNOTPSPKSLVEVNPKSEDAAGAGET 1066
      :      :      :      :      :      :      :      :      :
Db      726  VPGAPLAPLPINGRPFARKNLSL-----GSSSGDT 756

```

Search completed: April 26, 2002, 16:29:54
Job time: 609 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:21:42 ; Search time 153.5 seconds

(without alignments)
1315.022 Million cell updates/sec

Title: US-09-556-246-1_COPY_25_1404

Perfect score: 7410
Sequence: 1 QDLSSCAGRCGEYSRDATC.....ARATIRSGQTLSKMYNCP 1380

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organellar:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	7410	100.0	1404 4 Q92954	Q92954 homo sapien
2	7381	59.6	1404 4 Q9BX49	Q9BX49 homo sapien
3	3913	52.8	1054 11 Q9JW99	Q9JW99 mus musculu
4	1713.5	23.1	401 6 Q77765	Q77765 bos taurus
5	982.5	13.3	1079 5 Q9N4S7	Q9N4S7 caenorhabdi
6	946.5	12.8	1049 5 Q9I7S1	Q9I7S1 drosophila
7	902	12.2	1795 5 Q76894	Q76894 drosophila
8	863.5	11.7	1489 10 Q96449	Q96449 phytothphor
9	846	11.4	1274 5 Q20007	Q20007 caenorhabdi
10	816	11.0	1315 10 Q9SPM0	Q9SPM0 zea mays (m
11	812	11.0	2187 11 P70670	P70670 mus musculu
12	811.5	11.0	1480 10 Q9LIE8	Q9LIE8 arabidopsis
13	784	10.6	1188 10 Q41805	Q41805 zea mays (m
14	782.5	10.6	2112 5 Q9VEL9	Q9VEL9 drosophila
15	715	9.6	763 2 Q9N4S5	Q9N4S5 mycobacteri
16	709.5	9.6	6677 5 Q9N4S5	Q9N4S5 caenorhabdi
17	707.5	9.5	555 10 Q9FP06	Q9FP06 chlamydomon
18	703.5	9.5	4880 11 Q9JLT1	Q9JLT1 ratius norv
19	703.5	9.5	5085 11 Q9JKS6	Q9JKS6 ratius norv

20	695	9.4	1151 13 Q57580	Q57580 gallus gall
21	676.5	9.1	1229 5 Q94185	Q94185 caenorhabdi
22	674.5	9.1	2089 4 Q14676	Q14676 homo sapien
23	674.5	9.1	2768 5 Q9VC00	Q9VC00 drosophila
24	669	9.0	2284 5 Q9VPG1	Q9VPG1 drosophila
25	666.5	9.0	7962 4 Q10465	Q10465 homo sapien
26	665	9.0	4833 11 Q9OXY6	Q9OXY6 mus musculu
27	665	9.0	5038 11 Q9OXY7	Q9OXY7 mus musculu
28	663	8.9	3570 4 Q9S552	Q9S552 homo sapien
29	658.5	8.8	3507 5 Q9GUM7	Q9GUM7 caenorhabdi
30	653.5	8.8	1514 5 Q9GUM7	Q9GUM7 leishmania
31	643.5	8.7	990 13 Q91803	Q91803 xenopus lae
32	636	8.6	6632 5 Q17362	Q17362 caenorhabdi
33	633	8.5	489 10 Q41707	Q41707 vigna ungu
34	632	8.5	761 10 Q9Z010	Q9Z010 arabidopsis
35	631	8.5	2244 5 Q9N3Y8	Q9N3Y8 caenorhabdi
36	629	8.5	6642 5 Q01761	Q01761 caenorhabdi
37	628.5	8.5	1612 5 Q9VYQ2	Q9VYQ2 drosophila
38	627.5	8.5	971 5 Q9XV54	Q9XV54 caenorhabdi
39	622.5	8.4	839 2 Q9RX57	Q9RX57 delnococtus
40	607.5	8.2	801 5 Q23635	Q23635 caenorhabdi
41	607.5	8.2	924 12 Q93307	Q93307 epstein-bar
42	605	8.2	379 5 Q27929	Q27929 drosophila
43	600.5	8.1	1893 5 Q9NKC9	Q9NKC9 drosophila
44	598.5	8.1	409 10 Q9SBM1	Q9SBM1 volvox cart
45	591.5	8.0	2232 5 P91365	P91365 caenorhabdi

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	1404 AA.
Q92954	Q92954			
AC	Q92954	PRELIMINARY:	PRT:	1404 AA.
DI	01-FEB-1997 (TREMBLrel. 02, Created)			
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
DE	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
OS	MEGAKARYOCYTE STIMULATING FACTOR.			
OC	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OK	NCBI_TaxID=9606;			
OK	[1]			
RP	SEQUENCE FROM N.A.			
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
RA	Kellerer K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,			
RA	Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
RL	Blood 78:279-279(1991).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RA	Meisberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,			
RA	Fitzgerald M., Scalireto J., Kellerer K., Preissner K., Kriz R.,			
RL	Jacobs K., Turner K.,			
RL	(in) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,			
RL	Mosher D.F. (eds.);			
RL	Biology of vitronectins and their receptors., pp.45-52,			
RL	Elsevier Science Publishers B.V. (1993).			
RL	[3]			
RP	SEQUENCE FROM N.A.			
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
RA	Kellerer K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,			
RA	Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.			
RL	EMBL: U70136; AAB09089.1;			
DR	InterPro: IPR000585; Hemopexin.			
DR	InterPro: IPR001212; Somatomedin_B.			
DR	InterPro: IPR002400; GF_cysknot.			
DR	Pfam: PF00045; Somatomedin_B.			
DR	Pfam: PF01033; Somatomedin_B; 2.			
DR	PRINTS: PR00438; GRCYSKNOT.			

DR PRINTS; PRO0022; SOMATOMEDINB.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR SMART; SMO0120; HX; 2.
DR SMART; SMO0201; SO; 2.
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35FAF6 CRC64

Query Match	100.0%	Score 7410	DB 4	Length 1404
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1880; Conservative	0	Mismatches	0	Indels 0; Gaps 0

QY	1	ODLSSGRCGEGRSPATNCNVCNCHYKECCPDKRCVCIJELSGKGCFFSFEGRGC	60
Db	25	ODLSSGRCGEGRSPATNCNVCNCHYKECCPDKRCVCIJELSGKGCFFSFEGRGC	84
QY	61	DCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSKAPPPSGASOTIKSTTKRSKPPNNK	120
Db	85	DCDAQCKKYDKCCPDYSEFCAEVHNPTSPSSKAPPPSGASOTIKSTTKRSKPPNNK	144
QY	121	TKKYIESEITEEHSVEENQSSSSSSSSSTIMKIKSSKSAANRELQKKLYKDK	180
Db	145	TKKYIESEITEEHSVEENQSSSSSSSSSTIMKIKSSKSAANRELQKKLYKDK	204
QY	181	KNRTKKKPPKPVVDAGSGLDNGDKVTPDTSTTOHKNVSTSPKITTAAPINBRSL	240
Db	205	KNRTKKKPPKPVVDAGSGLDNGDKVTPDTSTTOHKNVSTSPKITTAAPINBRSL	264
QY	241	PNSGTSKETSUYNKETVYETKTTTNTKQTSIDGKEKTSKETSQSIJEKTSAKDLATP	300
Db	265	PNSGTSKETSUYNKETVYETKTTTNTKQTSIDGKEKTSKETSQSIJEKTSAKDLATP	324
QY	301	SKVLAKPPKAETTKGPAITTPKEPTPTTKKEBASTTPKEPTPTTIKSAPTTKKEBAPT	360
Db	325	SKVLAKPPKAETTKGPAITTPKEPTPTTKKEBASTTPKEPTPTTIKSAPTTKKEBAPT	384
QY	361	TTKSAPTTPKKPAPTTTKKEBAPTTPKKBPATTTTKKEBAPTITKSAPTTKKEBAPTKKA	420
Db	385	TTKSAPTTPKKPAPTTTKKEBAPTTPKKBPATTTTKKEBAPTITKSAPTTKKEBAPTKKA	444
QY	421	PTTPKEBAPTTPKEBPTPTPKKEBAPTITTKKEBAPTTPKBPATPAKKBAPTTPKEBAPTTPK	480
Db	445	PTTPKEBAPTTPKEBPTPTPKKEBAPTITTKKEBAPTTPKBPATPAKKBAPTTPKEBAPTTPK	504
QY	481	BPAPTTTKBESPTTPKEBAPTITTSAPTTTKKEBAPTITTKSAPTTTKBESPTTTKEBAPT	540
Db	505	BPAPTTTKBESPTTPKEBAPTITTSAPTTTKKEBAPTITTKSAPTTTKBESPTTTKEBAPT	564
QY	541	PKBPAPTTPKKPAPTTPKKEBAPTTPKKBPATTTTKBPATPAKKEBAPTTPKETAPTTPKKL	600
Db	565	PKBPAPTTPKKPAPTTPKKEBAPTTPKKBPATTTTKBPATPAKKEBAPTTPKETAPTTPKKL	624
QY	601	TTTTPEKLAPTTPKEBAPTTPBEELAPTTPEEPTPTTBEBAPTTPKAAANNTKKEBAPT	660
Db	625	TTTTPEKLAPTTPKEBAPTTPBEELAPTTPEEPTPTTBEBAPTTPKAAANNTKKEBAPT	684
QY	661	PKBPAPTTPKKPAPTTPKKEBAPTTPKGAPTTTLKEBAPTTPKKAPAKELAPTTTKBPTST	720
Db	685	PKBPAPTTPKKPAPTTPKKEBAPTTPKGAPTTTLKEBAPTTPKKAPAKELAPTTTKBPTST	744
QY	721	TSDDKAPTTPKGTAATTPKEBAPTTPKBPATTPKGAPTTTLKEBAPTTPKKAPAKELAP	780
Db	745	TSDDKAPTTPKGTAATTPKEBAPTTPKBPATTPKGAPTTTLKEBAPTTPKKAPAKELAP	804
QY	781	TTTTKGPSTITSKAPATTPKETAPTTTKKEBAPTTPKKAPATTPPTPPTSEVSTPTTK	840
Db	805	TTTTKGPSTITSKAPATTPKETAPTTTKKEBAPTTPKKAPATTPPTPPTSEVSTPTTK	864
QY	841	EPPTTIHKSPDESTPLSASPTPKALENSPKAPGATTKTPAATPKEMTTTAKOKTTERDL	900
Db	865	EPPTTIHKSPDESTPLSASPTPKALENSPKAPGATTKTPAATPKEMTTTAKOKTTERDL	924
QY	901	RTTPETTTAAPKMKETATTEKTESKIATATQVISTITTODTTPKITTLLKTTTLAPK	960

D	b	925	RTPPETTTAAKMKTKETATTEETTESKITTATTIOTVSTTTQDITTPRKITTLKTTTLAPK	984
Q	y	961	VTTTKTKITTTTELTMNKPDEEAKPKDRTATSKATTPPKPPTKAPKKTPTSKKPTMPBRV	1020
D	b	985	VTTTKTKITTTTELTMNKPDEEAKPKDRTATSKATTPPKPPTKAPKKTPTSKKPTMPBRV	1044
Q	y	1021	KPKTPTPRKMTSTMBELNFTSKIAEAMQLTTRPNQTPNSKIVEVNPKSADGAGSET	1080
D	b	1045	KPKTPTPRKMTSTMBELNFTSKIAEAMQLTTRPNQTPNSKIVEVNPKSADGAGSET	1104
Q	y	1081	PHMLLRPHVMPREVTPEMDIPLRPVPRNQGIIITNPMLSDENICNGKRPVUGLITLNGTLVA	1140
D	b	1105	PHMLLRPHVMPREVTPEMDIPLRPVPRNQGIIITNPMLSDENICNGKRPVUGLITLNGTLVA	1164
Q	y	1141	FRGHYFMMLSDEPSPSPARITTEVWQIIPSPIDIVTFRCCKEKTFFKDSOYMFRTNDIK	1200
D	b	1165	FRGHYFMMLSDEPSPSPARITTEVWQIIPSPIDIVTFRCCKEKTFFKDSOYMFRTNDIK	1224
Q	y	1201	DAGYPRDIFPKGFGGLTGOIYAAALSTAKYKNMDESVYFFKRGSSIIOOYIYKQBPQOKCGR	1260
D	b	1225	DAGYPRDIFPKGFGGLTGOIYAAALSTAKYKNMDESVYFFKRGSSIIOOYIYKQBPQOKCGR	1284
Q	y	1261	RPALNTEVYEGEMTOVRRRRRFEAIIIGBSQTHTRIOYSPARLAYQDKGYLVHNEVKSILMR	1320
D	b	1285	RPALNTEVYEGEMTOVRRRRRFEAIIIGBSQTHTRIOYSPARLAYQDKGYLVHNEVKSILMR	1344
Q	y	1321	GLPWNVTSALSLPNIRKPDGYDYVAASKOYNYINDPSTARAIIITRSGOITLSKWYNCP	1380
D	b	1345	GLPWNVTSALSLPNIRKPDGYDYVAASKOYNYINDPSTARAIIITRSGOITLSKWYNCP	1404

RESULT	2	
Q9BX49		
ID	Q9BX49	PRELIMINARY; PRT: 1404 AA.
AC	Q9BX49;	
DT	01-JUN-2001 (TrEMBLrel. 17, Created)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	
DE	BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR).	
GN	BG174L6.2.	
OS	Homo sapiens (human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RX	[1]	
RP	SEQUENCE FROM N.A.	
RA	Wray P.;	
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AL133553; CAC36090.1; -	
SQ	SEQUENCE 1404 AA: 151076 MW: 782A11746B3FDEE5 CRC64:	

Query Match	99.6%;	Score 7381;	DB 4;	Length 1404;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 1376;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0.

Qy	1	ODLSSCARCGGYSRDAITCNCDVNCQHYMCCPDFFRVCAETLSCSGRPFESEREREC	60
Db	25	ODLSSCARCGGYSRDAITCNCDVNCQHYMCCPDFFRVCAETLSCSGRPFESEREREC	84
Qy	61	DCDAQCKYKDKCCPDYESFCAEVHNPTSPSSKAPPSGASOTIKSTYTRSPKPPMKK	120
Db	85	DCDAQCKYKDKCCPDYESFCAEVHNPTSPSSKAPPSGASOTIKSTYTRSPKPPMKK	144
Qy	121	TKKVIESEETIEHVSVENOESSSSSSSSSTIKIKSSKNSAARELOKIKLYVNDK	180
Db	145	TKKVIESEETIEHVSVENOESSSSSSSSSTIKIKSSKNSAARELOKIKLYVNDK	204
Qy	181	KKRTKKKPPRPVYVDEGSGLDNCDFFVTTPDSTIOHKNVSPKITYAKPINPAPSL	240
Db	205	KKRTKKKPPRPVYVDEGSGLDNCDFFVTTPDSTIOHKNVSPKITYAKPINPAPSL	264

```
QY 241 PENSISKETSLVANKETVETKETTNNKQSTDGKCKTTSKAKETOSIEKTSKDLAPT 300
DB 265 PPNSDTSKETSLSLVNKKETVETKETTNNKQSTDGKCKTTSKAKETOSIEKTSKDLAPT 324
QY 301 SKVLAPTPKAEETTTGPAITTPKEPPTTPKBPASTTPKEPPTTTKSAPTTPKEPAPT 360
DB 325 SKVLAPTPKAEETTTGPAITTPKEPPTTPKBPASTTPKEPPTTTKSAPTTPKEPAPT 384
QY 361 TTYSAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKPA 420
DB 385 TTYSAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKPA 444
QY 421 PTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAPATTPKBPAPTTPK 480
DB 445 PTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAPATTPKBPAPTTPK 504
QY 481 EPAPTTPKESPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP 540
DB 505 EPAPTTPKESPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP 564
QY 541 PKBPAPTTPKBPAPTTPKEPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 600
DB 565 PKBPAPTTPKBPAPTTPKEPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 624
QY 601 TPPTPEKLAFTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAAAPTTPKEPAPT 660
DB 625 TPPTPEKLAFTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAAAPTTPKEPAPT 684
QY 661 PKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 720
DB 685 PKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 744
QY 721 TSDKPAFTTPKGAFTTPKEPAPTTPKEPAPTTPKGAFTTPKEPAPTTPKGAFTTPK 780
DB 745 TSDKPAFTTPKGAFTTPKEPAPTTPKEPAPTTPKGAFTTPKEPAPTTPKGAFTTPK 804
QY 781 TTTKGGSTSTTSOKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 840
DB 805 TTTKGGSTSTTSOKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 864
QY 841 EPTTIKHSDESPTELKSAETPKALNSPKEPVPPTTKPAAKPPMTTAKKTEEROL 900
DB 865 EPTTIKHSDESPTELKSAETPKALNSPKEPVPPTTKPAAKPPMTTAKKTEEROL 924
QY 901 RTTPETTTAARPKMTKEATTTKTESKITATTTQVSTTTODTTPFKITTLTKTTTLAR 960
DB 925 RTTPETTTAARPKMTKEATTTKTESKITATTTQVSTTTODTTPFKITTLTKTTTLAR 984
QY 961 VTTTKKTITTTTEIMNKPETAKPKDRATNSKATTPKQAPTKAPKPTSTKKKTPRVR 1020
DB 985 VTTTKKTITTTTEIMNKPETAKPKDRATNSKATTPKQAPTKAPKPTSTKKKTPRVR 1044
QY 1021 KKTPTTPPKMTSTMBELNPTSRILAEAMLOTTTRPQOTNSKLVENPNSSEDDGAGEET 1080
DB 1045 KKTPTTPPKMTSTMBELNPTSRILAEAMLOTTTRPQOTNSKLVENPNSSEDDGAGEET 1104
QY 1081 PHMLLRPHVFMPEVETDMDYLRPRVNOGIIINPMLSEDTNICKGKPVDDGLTTLRNTTLVA 1140
DB 1105 PHMLLRPHVFMPEVETDMDYLRPRVNOGIIINPMLSEDTNICKGKPVDDGLTTLRNTTLVA 1164
QY 1141 PFGHVFMMLSPPSPSPARITEVWGIISPIDIVFTRCCEGTFFFKKSQVWRFNTNDIK 1200
DB 1165 PFGHVFMMLSPPSPSPARITEVWGIISPIDIVFTRCCEGTFFFKKSQVWRFNTNDIK 1224
QY 1201 DAGYPAKIFKGFGLGQIVAAALSTAKYKNMPESVYFFKRGSGIOQYITKQBPVOCSPGR 1260
DB 1225 DAGYPAKIFKGFGLGQIVAAALSTAKYKNMPESVYFFKRGSGIOQYITKQBPVOCSPGR 1284
QY 1261 RALANLTVYGEAMTOVARRRFPERAIGDSQHTTIRIOYSPARLAYQDKGVLJHNEKYSILMR 1320
DB 1285 RALANLTVYGEAMTOVARRRFPERAIGDSQHTTIRIOYSPARLAYQDKGVLJHNEKYSILMR 1344
QY 1321 GLPNVTVSALSLPNIRKPDGYDYVAFSKQOYINIDVPSTATAITTRSGOTLSKVMYNCP 1380
```

```
DB 1345 GLPNVTVSALSLPNIRKPDGYDYVAFSKQOYINIDVPSTATAITTRSGOTLSKVMYNCP 1404

RESULT 3
ID 09Jm99 PRELIMINARY; PRT; 1054 AA.
AC 09Jm99:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE mRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN PRG4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikegawa S., Nakamura Y.;
RT "a novel mouse gene highly homologous to a human gene encoding
RT megakaryocyte stimulating factor precursor and cartilage superficial
RT zone protein.";
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB034730; BAA92310.1; -.
DR MGD: MGI:1891344; Prg4.
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR002965; P_rich_extensn.
DR InterPro: IPR001212; Somatomedin_B.
DR Pfam: PF01033; Somatomedin_B; 2.
DR PRINTS: PR01217; PRICHEXTENS.
DR PRINTS: PR00022; SOMATOMEDINB.
DR SMART: SM00120; HX; 2.
DR SMART: SM00201; SO; 2.
DR PROSITE: PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE: PS00524; SOMATOMEDIN_B; 2.
DR SEQUENCE 1054 AA; 115991 MW; 4FC6BFA42283235 CRC64;

Query Match 52.8%; Score 3913; DB 11; Length 1054;
Best Local Similarity 57.1%; Pred. No. 4,6e-245;
Matches 795; Conservative 57; Mismatches 166; Indels 374; Gaps 25;

QY 1 ODLSGAGCGEGYSRDATCNDYNCQHYMECCPDRCRYCTAELSGKGFESFEGREGC 60
DB 25 QDLSGAGCGEGYSRDATCNDYNCQHYMECCPDRCRYCTAELSGKGFESFEGREGC 84
QY 61 DCDAGCKKYDKCCPDYSEFCAEVHNPTSPSSKKAPPGASQTIKSTTKRSPKPPNKK 120
DB 85 DCDAGCKKYDKCCPDYSEFCAEVHNPTSPSSKKAPPGASQTIKSTTKRSPKPPNKK 142
QY 121 TKKVISEELTEHVSSENOESSSSSSSSSTTWIKIKSKNSAANRELQKKLVKDK 180
DB 143 TKKVISEELTEHVSSENOESSSSSSSSSTTWIKIKSKNSAANRELQKKLVKDK 199
QY 181 KNRKTKKPPKPPVYDEAGSGLDNGDEKFT--TPDSTQNHKKVSVSPKITTAKPINPRP 238
DB 200 KNRKTKKPPKPPVYDEAGSGLDNGDEKFTLPPEPDPTTPHKSIVATSPYTTAAKVPYTPR 259
QY 239 SLPPNSDTSKETSLVANKETVETKETTNNKQSTDGKCKTTSKAKETOSIEKTSKDLA 298
DB 260 SLPPNSDTSKETSLVANKETVETKETTNNKQSTDGKCKTTSKAKETOSIEKTSKDLA 317
QY 299 PTKSVLAKTTPKAEETTTGPAITTPKEPPTTPKBPASTTPKEPPTTTKSAPTTPKEPA 358
DB 318 PTKSVLAKTTPKAEETTTGPAITTPKEPPTTPKBPASTTPKEPPTTTKSAPTTPKEPA 336
QY 359 PTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 418
DB 337 PTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPK 381
QY 419 PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAPATTPKBPAPTTP 478
```

[illegible]

```

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SUPERFICIAL_ZONE_PROTEIN (FRAGMENT).
OS Bos taurus (bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ARTICULAR CARTILAGE;
RA Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,
RA Ayselote M.B.;
RT *Immunodetection and Partial cDNA Sequence of the Proteoglycan,
RT Superficial Zone Protein, Synthesized by Cells Lining Synovial
RT Joints."
RL J. Orthop. Res. 0:0-0(1998).
DR EMBL; AF056218; AAD13404.1; -.
DR InterPro; IPR00585; Hemopexin.
DR Pfam; PF00045; hemopexin; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR SMART; SM00120; HX; 2.
FT NON_TER 1
SQ SEQUENCE 401 AA; 44952 MW; 86147CC9AFBB73D7 CRC64;

```

RESULT	4	
077765		
ID	077765	PRELIMINARY; PRT; 401 AA
AC	077765;	
DT	01-NOV-1998	(TREMBLrel. 08, Created)

Query Match	23.1%	Score 1713.5	DB 6	Length 401
Best Local Similarity	80.2%	Pred. No. 2.7e-103		
Matches 320	Conservative	24	Mismatches 52	Indels 3
QY 983	PKDRATNSKATTPKQKPTKAPKKTSTYKPKTKMTKRVKRPKTTPTPRK-MTSTMPELNPT	1041		
DB 5	PGGRATNSQVTPKPKQKPKAPKPKPTSTKKPKPT-PRVRRPKTTTPPKTTTSAMPEPTPT	63		
QY 1042	SRIAEAMLOTTRPQTPKNSKLVENVKPSKEDAGAEGETPHMLLRPHVFMPEVTEDMDYL	1101		
DB 64	S-LPEAMLOTTRPPTPTPNSKELIDVNSEMDGDAAGEKPHMIFRPVLTPLVIGTEII	122		
QY 1102	PRVPGMIILNPMLSDEFINICKGKPYDGLITLRNGTLVAFRCHITWMI-SPESSPSPART	1161		
DB 123	VGPSGSGEINPMFSDETINCGRPYDGLITLRNGTLVAFRCHITWMLLTPFPPEPPRI	182		
QY 1162	TEWNGIIPSIDVPEVPCNCEGKTFPKDSQVRFNTDIDACVYPIFKFGFGLIGUUA	1221		
DB 183	TEWNGIIPSIDVPTFRNCEGKTFPKDSQVRFNTDIDACVYPIKSGFGGLNGKIVA	242		
QY 1222	ALSTAKYKMWPSVYFFKRGGSIOQYIVKQEPVQKCGRRPALNPVYGEMTQVRRRPE	1281		
DB 243	ALSIADQKSRPSVYFFKRGGSVOQYIVKQEPVQKCGRRPALNPVYGETAQVRRRPE	302		
QY 1282	RAIGPSQVTHIRIYSPALAYODGCVLHNEKVSILTMGSLPNVYTSALSLNIRKPDQY	1341		
DB 303	RAIGPSQVTHIRIYHPVVPYODGKGLHNEKVSILTMGSLPNVYTSALSLNIRKPDQY	362		
QY 1342	DYAFESKDQYINIDVPSRTARAITTRSGOTLSKVMYNCP	1380		
DB 363	DYALSKDQYINIDVPSRTARAITTRSGOTLSNTWYNCP	401		
RESULT 5				
Q9N4S7	PRELIMINARY;	PRT; 1079 AA.		
AC Q9N4S7				
DT 01-OCT-2000	(TREMBLrel. 15, Created)			
DT 01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT 01-JUN-2001	(TREMBLrel. 17, Last annotation update)			
DE Y5IB11A.1	PROTEIN.			
GN Y5IB11A.1				
OS Caenorhabditis elegans.				
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;				
OC Rhabditidae; Peloderinae; Caenorhabditis.				
OX NCBI_TaxID=6239;				
RN [1]				

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Cotton M.;
RT "The sequence of C. elegans cosmid Y51B11A."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL EMBL: ACC06797; AAF60743.1;
DR InterPro: IPR002965; P_rich_extensn.
DR PRINTS: PR01217; PRICHEXTENS.
SQ SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;

Query Match 13.38; Score 982.5; DB 5; Length 1079;
Best Local Similarity 29.68; Pred. No. 1.4e-55;
Matches 332; Conservative 131; Mismatches 481; Indels 177; Gaps 44;

OY 171 OKLKVKNKKRKKKPKPVNDEAGSLDNGDEKV-----TTDDSTTQ-HNKV 222
DB 27 OKLELADCANQPTHTMLP-----STLFSVMEPTSLVLSSTPSSSTPKET 77
OY 223 STSPKITTAKPVPNSDTSKETSIVNKEET-VERKEETITNKQSTGKEKT 281
DB 78 TTAPEETSTEP-----PSSSTTPVQTTTTTAPETTSPEASSSTPVQTT-----TT 124
OY 282 SAKETOSIEKTSKOLAPSKVLAKPTKAET-TKGP-ALITPKPPTTKEPASTTP 339
DB 125 TAPETSTEPSS-----TSVQTTTTTAPETTSPEASSSTPVQTTTTTAPETSTEP 180
OY 340 KEPTPTTISAPPTPEPAPTTTSAPPTTKEPAPTTKEPAPTTKEPAPTTKEPAPT 399
DB 181 PSSSTTPVQTTTTTAPETTSPEASSSTPVQTTTTTAPETTSPEASSSTPVQTT 236
OY 400 TTKSATTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPA 459
DB 237 TTTTAPETTSPEASSSTTPVQTTTTTAPETTSPEASSSTTPVQTTTTTAPETTSPE 295
OY 460 PTAPKPAPTTKEPAPTTKEP-----APTTKEPSTTPKEPAPTTTS 505
DB 296 PSSSTTPVQTTTTTAPETTSPEASSSTTPVQTTTTTAPETTSPEASSSTTPVQTTT 355
OY 506 APPTTK--EPADPTTKSAPPTTKEPSTTPKEP-APTTPEKAPPTTKEPAPTTKEPA 562
DB 356 ABEETRTPESSSTTPVQTTTTTAPETTSPEASSSTTPVQTTTTTAPETTSPE 411
OY 563 TTPKPEAPPTTKEPAPTTKEP-APTTPEKAPPTTKEPAPTTKEPAPTTKEPAPT 620
DB 412 SSSTTPVQTTTTTAPETTSPEASSSTTPVQTTTTTAPETTSPEASSSTTPVQTT 467
OY 621 PELAPPTPEEPPTPEPAPTTKEPAAPPTTKEP-APTTPEKAPPTTKEPAPTTKEP 678
DB 468 TTTAPETTSPEASSSTTPVQTTTTTAPETTSPEASSSTTPVQTTTTTAPETTSPE 525
OY 679 ETAPPTPGTAPTTKEPAPTTKEPAKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 737
DB 526 PSSSTTPVQTTTTTAPETTSPEASSSTTPVQTTTTTAPETTSPEASSSTTPVQTTT 585
OY 738 PKEPAPTTKEP-APTTPEKAPPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEP 794
DB 586 APE-----TSTEPSSSTTPVQTTTTTAPETTSPEASSSTTPVQTTTTTAPETTSPE 642
OY 795 APPTPEKAPPTTKEPAPTTKEP-----PAPPTPEPTPTT-----EVSPTTTTKE 841

DB 643 SSTTPVQTTTTTAPETTSPEASSSTTPVQTTTTTAPETTSPEASSSTTPVQTTTAP 702
OY 842 PTTHKSPDESPETLAEPT--PKALENSPKKEGVTTTTPATKREMTTAKKKTEND 899
DB 703 ETSTPESSSTTPVQTTTTTAPETTSPEASSSTTPVQTTTTTAPETTSPEASSSTTP 762
OY 900 LRTT-----PETTAAKMKETATTEKTESKITATTTQVSTTTQDTTPTKTTKT 955
DB 763 VQTTTTTAPETTSPEASSSTTPVQTTTTTAPETTSPEASSSTTPVQTTTAPETTT 820
OY 956 TLAPKTTTNTTTTTEINMKPEETR-KPKDAPNLSKATTPPKQAPKAPKPTSTKPK 1014
DB 821 STEPSSSTTPVQTTT--TTAPETTSPEASSSTTPVQTTTTTAPETTSPEASSSTTPV 878
OY 1015 TMRVKKPTTPTPKMTSTPELNPSTIAAMQTTTRPQNTSKIVENPKSEDA 1074
DB 879 QTTTITAPETTSPEASSSTTPVQTTTTTAPETTTSTEP--PSSSTTPVQTTTTTAP 932
OY 1075 GABG-ETPMILRPVMEVTP---DMDYLPV-----PNOGIINPMISDEFN 1120
DB 933 ETTSTEPSSSTTPVQTTTTTAPCDCLSYIDRVYVPTTEWENKROIITQSYSPRRT 992
OY 1121 -----TNGKRPVGLTTRNGTIVARGHFMLS-----PSPSPS 1156
DB 993 AFSFSTPDIGCTALILCTITSEGISNL-NATL-----TGLSDGSSIDLFF----- 1040
OY 1157 PARRTIEWGLPSIDTFTTRCNCBKTFFFKDSQYMRFTN 1197
DB 1041 -----YNPVTGLE-IMPEINCEGNK-----WSYNN 1064

RESULT 6
O917S1 PRELIMINARY; PRT; 1049 AA.
AC 0917S1;
DT 01-MAR-2001 (Tremblrel. 16, created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE CG5228 PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sulten G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Milos G.L.G.,
RA April J.F., Ashgiani A., An H.-J., Andrews-Plamkoc H., Balaban D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Betman B.P., Bhandari D., Bolshakov S.,
RA Borok D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahiker C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Giodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostlin D., Houston K.A., Howland T.J., Mel M.-H., Ibegyan C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shier B.C., Siden-Kiamos I., Simpson M., Skipski M.P., Smith T.,
 RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 RA Svetskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195 (2000).
 DR EMBL: AE003495; AAG22353.1;
 DR Flybase: Fggn0030561; CG5228.
 SEQ SEQUENCE 1049 AA; 107278 MW; 954DD629E7619671 CRC64;

Query Match 12.8%; Score 946.5; DB 5; Length 1049;
 Best Local Similarity 34.4%; Pred. No. 2.9e-53;
 Matches 376; Conservative 69; Mismatches 480; Indels 169; Gaps 56;

QY 82 EVHNPTSPSSSKRAPPSSGASOTIKSTKRSKPKPKKKTKVIESEETEEHSUSENO 141
 DB 33 EIGTAAKPTTLKPTEGTSAPTLKPTEGTSAPTLKPT-----EGTAAKPTTLKPT 86
 QY 142 SSSSSSSSSSSSTIWKIKSSKSNANRELQKLVKNKKNKTKKPKPPVDEAGSG 201
 DB 87 GTTAAKPTT-----LKPEGTSA-----KPTTLKPT-----GTTAAKPTTLKPT 128
 QY 202 LDWGDFFVTTPTDS--TTOHNKYSTSPKITTAAPINPROSLPNSDTSKETSITVKNKT 259
 DB 129 AKPTTLKPTEGTAAKPTTLKPTEGTSAPTLKPTEGTSAPTLKPTEGTSAPTLKPT 185
 QY 260 VETKETTITTKQT--STDGKERTSAKET--QSIETKSAK--DLAPPSKVLAKPT--PK 310
 DB 186 LKPEGTSAKPTTLKPTEG--TSAPTLKPTEGTAAKPTTLKPTEGTSAPTLKPT 241
 QY 311 AETTKGPAALTTPKEPTTPPKPEASTPKPT--PTTIKSAPTTPKEPA--PTTTS 364
 DB 242 EGTAAK--PTTLKPTEGTAAKPTTLKPTEGTSAPTLKPTEGTAAKPTTLKPTEGTS 297
 QY 365 APTPKPEAPPTTKEPA--PTPKPEAPPTTKEPAPTTTSAPPTPKPEAPPTPKAP 421
 DB 298 AKPTTLKPTEGTAAKPTTLKPTEGTSAPTLKPTEGTAAK--PTTLK--PTEGTSAP 352
 QY 422 TTPKEAPPTPKPT--PTPKPEAPPTTKEPAPTTKEPA--PTTAKKPEAPPTPKPEAP 476
 DB 353 TTKLPDTGTTAAKPTTLKPTEGTSAPTLKPTEGTAAKPTTLKPTEGTSAPTLKPT 412
 QY 477 TTPKEAPPTTKEPSPTPKPEAPPTTTSAPTTTKEPA--PTTTSAPPTPKPEPSPTT 533
 DB 413 TTAK--PTTLKPTEGTAAK--PTTLKPTEGTAAKPTTLKPTEGTSAPTLKPTGTT 466
 QY 534 KEPAPTPKPE--PAPTPKPEAPPTTKEPAPTTKEPAPTTTSAPPTPKPEAPPTPK 587
 DB 467 AKPTTLKPTEGTAAKPTTLKPTEGTSAPTLKPTEGTAAKPTTLKPTEGTAAKPT 525
 QY 588 TPKETAPPT--PKKLTPTTPPKLAPTT--PEKPAPTTPEELAPTT--TPEEPTTPPEEPA 641
 DB 526 TLKPTKGTAAKPTTLKPTEGTSAPTLKPTEGTAAKPTTLKPTEGTSAPTLKPT 585
 QY 642 PTTPKAAAPNTPKPEAPPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTT 701
 DB 586 TAKPTTLKPTEGTSAPPTTLK--PTEGTSAPSTTLKPTEGTAAK--PTTLK--PTEG 636
 QY 702 KKPAPELAPTT-----TTKEPTSTSDKPAPTTPKGTATPTTKEPAPTTKEPAPTT 754
 DB 637 TSAOFTTLKPTEGTAAKPTTLKPTEGTSAPTLKPTEGTAAK--PTTLKPTDGTGA 691
 QY 755 KGTA-----PTTLKPEAPPTPK--KPAPELAPTTTGGPTSTSDKPAPTTPK 800

DB 692 KPTTLNPEGTSAKPTTLKPTEGTAAKPTTLKPTEGTSAPTLKPTEGTAAKPTTLKPT 751
 QY 801 ETAPPTKEPAPTT--PKKPAPTTPEPTPTTSEVSPPTTK-----EPTTIKSPDE 851
 DB 752 EGTATTA--PTTLKPTEGTSAPTLKPTEGTAAKPTTLKPTEGTSAPPTTL--KPT 806
 QY 852 STEPLSAPPT--KALENSPKRGV--PTTKT--PAATKREMTTAAKDKTERDRLTPE 905
 DB 807 RT--SAOPTTLKPTEGTAAKPTTLKPTEGTSAPPTTLKPTEGTAAKPTTLKPT 858
 QY 906 TTPAAPTKEPTATTTETTESKITATTTTODTTPPKITTLKTTTLAKPVYTTK 965
 DB 859 GTSAPKPTTLKPTEGTAAKPTTLKPTEGTSAPTLKPTEGTAAKPTTLKPTDGTGA 914
 QY 966 K--TTTTTEINNKPEETAKPKDRATNSKATTPKQKPTAKP--KPT--STKKKMPR 1018
 DB 915 KPTTLKPTEGTSAPPTTLKPTPE--GTTAAKPTTLKPTEGTSAPPTTLKPTEGTAAKPTTLKPT 973
 QY 1019 VRKPTTPT--PRKMTSTMP--ELNPTSRIAEAMLOTTTTPRQTPNSKIIVEVNEKSDAG 1074
 DB 974 TEGPSAKPTTLKPTERTISAOPTTLKPTERTISAQ--PTTLKPTEGTAAKPTTLKPTKGTSG 1031
 QY 1075 GAEG--ETPHMLRP 1087
 DB 1032 QANNETKRKKERR 1045

RESULT 7
 076894 PRELIMINARY; PRT; 1795 AA.
 AC 076894;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE EG:5667.1 OR CG14796.
 GN EG:5667.1 OR CG14796.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RC MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Burton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Abmayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fioder C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Foster A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Kelum K.A.,
 RA Kimmel B.R., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spirding A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskaya R., Tector C., Turner R., Venter F., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhou Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 RP [2]
 RP SEQUENCE FROM N.A.
 RA Cadieu E., Dreano S., Lelaure V., Mottier S., Galibert F.,
 RT "Sequencing the distal X chromosome of *Drosophila melanogaster*."
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RP [3]
 RP SEQUENCE FROM N.A.
 RA Benos P.,
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF003421; AAF45644.1; -
 DR EMBL: AL031028; CAI19845.2; -
 DR Flybase: FBgn0025390; EG:5667.1.
 DR InterPro: IPR002557; Chitin_binding.
 DR InterPro: IPR002965; P_rich_extensn.
 DR Pfam: PF01607; Chitin_bind_2; 2.
 DR PRINTS: PR01217; PRICHEXTNSN.
 DR SMART: SM00494; ChEBD2; 2.
 SO SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match 12.28; Score 902; DB 5; Length 1795;
 Best Local Similarity 25.98; Pred. No. 3.8e-50;
 Matches 330; Conservative 163; Mismatches 537; Indels 246; Gaps 43;
 QY 1 QDLSSCAGRC-----GEG-YSRDACCNDYNC-----QHVECCP--- 34
 DB 124 ESFNASSRCQHVPRHPRHRCQCKEGFRPHDCKYVYRCDKRNRTQPMWLFACPAGT 183
 QY 35 ---DFKVC-----TAELSCNG-----RCFE--SPERGECDCOACKKYD 70
 DB 184 IFSEVFERKCLPGDCPSFEISDSGYIPQNCCLKFPCEAEGFRSTDDALYYTCRLD 243
 QY 71 K-----CCPDYSCFAE-----VHNPTSP---PSSKAPP-- 97
 DB 244 SGTYLQTRFCQGSNSFDLERKLCRPRESEVDCDFVPGVQVYPAQVYPPYAPAPLY 303
 QY 98 -----PSGASQ---TIKSTKRSPKPPNKKTKKVISESEITEHSGVSEMOSSSSSS 149
 DB 304 EEDDYDTGAREQOPALKEKQLQVAAGEFEKPSLVNVLQTTTLEPSTAIHKIPAYPSYPS 363
 QY 150 SSSSTIKIKSKNSAANRELQ-----KLKVKDN----- 179
 DB 364 YETSS--HHNGKRAEMENLEKEGVPRKLISENIIVIOPEPATAATREPLINDINKY 421
 QY 180 -----KKNRKKKPPKPPRVNDEAGSLDNGDFKVTTPDSTQHNKVSFKITTA 231
 DB 422 YKRYVTGTDKNVTEAPKSP-----KQHLSENIIVLPEYTT--TTTTKPVYVLC 474
 QY 232 KPIPNRPSLPNSDT-----SKETSLVKNKETTVEKTKETTTNKOTSTGKEKT 280
 DB 475 PTTSPDTPPKPSTTVAVTKSPKISSTEQHSITTAKTITTKRPTTVTEKTSANKEPKT 534
 QY 281 TSAKETOSIEKTSADKLAPTSVKLAPPKAKETTKGALTPPKKEPTTPPKKEPASTPK 340
 DB 535 TVVTTTQKRSTTNTSPTDKTTITSTLSKPTTTPPTSTPTSTPTST-----TPPS 588
 QY 341 EPPPTTKAPPTPKKAPATTKSAPTTKKEPAPTTK--EPAPTTKKEPAPTTKKEPAPT 399
 DB 589 TTTPTST-----TTPSTTTTSTTTTVKVSHPBRSTSQKTTASTTTTK-----TTTSPKTT 639
 QY 400 TTKSATPTTKEPAPTTPKKAPATTKEPAPTTKKEPTPTTTPKEPAPTTKKEPAPTTK 459

DB 640 KTTDIPSTYSKLSSTTQKTTTTTHKFTAAVTTSEKPKTTKSTSVSTTTKSTESSPK 699
 QY 460 PTAKKAPATTTPKKEPAPTTKKEPAPTTTKEPSPPTTKEPAPTTTKSAPTTKKEPAPTTK 519
 DB 700 PTTSGAPTTTPKSTSTTTTPPTTKVTVTQTITTTPLRSTETVSTQPPYTTTPQPTTTT 759
 QY 520 SAPTTKPEPSTTPKKEPAPTTKKEPAPTTK-----KRPATTPKKEPAPTT----- 562
 DB 760 TLVTPTKSTTTTTEKPIISSKPTTTTQKTTSTANNTKVAITTOKEETTPQOSTVIT 819
 QY 563 -----TTPKEPAPTTTPKAPAPKKEPAPTPK--ETAPTPPKLLPTTPEKLAETTPK 616
 DB 820 FTKRTTNNPEPSTKPTSTPTKPTSTTTPKSTVAASSTKETTISPCKTTEKSTENPT 879
 QY 617 -----APTPPELAETTPPEEPPTTPEEPAPTTTKEAPAPTTKKAAPNPKKEPAPTT 667
 DB 880 TNSVKTALISSQKATSTSEPTKTT-QNITTTTPKPTTLKSTQEAATSTQKSTVIT 938
 QY 668 TPKEPAPTTKKEPAPTTPKGAPATTLKEPAPTTPKKPAKELAPTTTKEPT--STSDPK 725
 DB 939 TTKKATSESLTLTSTEEPNTTPKPLRTTPTTSTVATTRITTTTISSESTETTOKP 998
 QY 726 APPTPKGIAPTTPKKEPAPTTPKKEPAPTTKGPATTLKKEPAPTTKKEPAPTKELAP--TTTK 784
 DB 999 KSTPTSTTTTTPKVTTVVSTQNPPTTTSKSTVVI-----TTP-NPSPSTQRPPTTTR 1052
 QY 785 GPTSTSDKRAPTTPKKEPAPTTPKKEPAPTTPKKAPATTEPPTTSEVSTPTT--TKE-- 841
 DB 1053 QPSTIASSTSIGTIRLPTTTPNQNSTSDLTLYTRPCPPDPDSTDKNTNFACTQELQ 1112
 QY 842 -----PTTIKSPDESTPELSAEPTRPALENSPKREPVPRTKTPAATK 884
 DB 1113 QVNLELSPQKOEOPHTHTHTALGSRNLTGCOEVPVMDAPSAEBSGQATTAKA 1172
 QY 885 PEWTTAKDKTER--DLRTTPETTTAPKMTKEATTEKTTTSKITATTTOVTSTTT 941
 DB 1173 PIMSTIAAHLLQKLFHITSTTPPSREHAP--TONPSSOPSSOSQR--GVYTAQARHNL 1229
 QY 942 QDTPFKITTLKT--TLAPKVTTKTITTTTEIMKPEETAPKRAATNSKATTPK 997
 DB 1230 AVSKPEIASLRLSLQOLASTQKRSIPKTLVHTNHTKEPD--SEVYSERSEQVTDSDN 1288
 QY 998 OKPTKA-PKKPTSTKPKPTMPVRKPKTTPPKKTTSTMPELNPTSLAEMLQTTTTPN 1056
 DB 1289 EYLDKTOPRAMSTVAVALPVPSTTEREPQK--TSSSP--SPT--KATSSPTTQPI 1341
 QY 1057 QTPNSKLVEVNPESD 1072
 DB 1342 ETTTGDEYDSSGSSD 1357
 RESULT 8
 096449 PRELIMINARY; PRT; 1489 AA.
 AC 096449;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DE 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
 OS CA90.
 OS Phytophthora infestans (Potato late blight fungus).
 OC Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
 OC Phytophthora.
 OX NCBI_Taxid=4787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RACE 1-11;
 RA Goerinaard B.;
 RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF061185; AAC72308.1; -
 SO SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;

	Query Match	11.7%	Score 863.5:	DB 10:	Length 1489;	
	Best Local Similarity	31.7%	Pred. No. 9.6e-48:			
	Matches 368;	Conservative 55;	Mismatches 509;	Indels 229;	Gaps 51.	
OY	121	TKKVESEETIEHSHENSENOESSSSSSSSSSSTIMWIKSKSNANRELOKIKAYDKN	180			
Db	282	TPIVGYSTEETEONHTGGEPEPDEAPTECTTY - VPREEYAAAPSE-----DTIYAP	334			
OY	181	KNRKKKKPPKPVVDADAGSLDNDDFKYTIDTJST-----QHNWYSUSPKITTKAKPI	234			
Db	335	REVTPYAPTEKPYDVBETYYVTVEESTYAPTKESTNAIPERKHIAHIEPCDETWTMAYPT	394			
OY	235	NPRRSLP-----PNSDT-----SKETSINKEETIVEKETTT---TNKOT	272			
Db	395	EETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAP	454			
OY	273	STDCEKETSASKEOSIOLEKTASDLAPISVKYLAKPIPAEITTKCPALTTKEPTPTPK	332			
Db	455	YAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAP	511			
OY	333	EPASTPEKEPPTTIKSAPTTKEPAPPT--TKSAPTTKEPAPPT--TKREPAPTTKEPA	389			
Db	512	APAETPEPEPEEL-TYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPAEETIYAPEPT	570			
OY	390	PTTKREPAPTTKSAPT-----TKREPAPTTPKRAPTTKEKPAPTTPKRPPTT	439			
Db	571	EETIYAPTEETT-YAPTEETIYAPIEBITTYAPTEETIYAPAEETIYAPEETIYAPTEET	629			
OY	440	PKBP-----AP-----TKREPAPTTKEP-----APTA PKKPAPTTKEPAP	476			
Db	630	TYAPTEETIYASTEETIYAPTEETIYAPAEETIYAPEETIYAPTEETIYAPTEETIYAP	689			
OY	477	TTTTKPADPT--TKEPSPTTKPEP-----APT--TKSAPT-----TKKBP-----	513			
Db	690	TEETIYAPTEETIYAPAEETIYAPEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAP	749			
OY	514	--APT--TKKSAPT--TKPEPSPTTKEP-----APTTKEPAPTTPKKPAPTTKEPAP	562			
Db	750	TYAPTEATTYATEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAP	809			
OY	563	T--TKREPAPTTKKRAPAPAKEPAPTTKEPATPTTKLPTTPPKLAPPTPEKAPPT	620			
Db	810	TEETPEPEPTTEETIYAPTEETPEPEPTET--TYTPEETIYAPTEETIYAPTEETIYAPTE	867			
OY	621	PERIAPT--TPEEPPTPEEP-----APT-----TKAAMPYTPKEPAPTTPK	662			
Db	868	EYTAPEETPEPEPEETIYAPTKETIYAPTEETIYASTEETIYAPTEETIYAPAEETIYAP	927			
OY	713	-TTKEPTSTSOKRAPTTPKGTAPTTKEPAPTTKEP-----APTTPKGTAPPT-----	760			
Db	988	EYTAPEETEMVAPIEBITTYAPTEETIYAPAEETIYAPEETIYAPTEETIYAPTEETIYAP	1047			
OY	761	-----TLKEPAPTTPKKP-----APKE--IAPT--TKGPSTTSOKPAP	796			
Db	1048	ASTBEETIYAPTEETIYAPAEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTE	1107			
OY	797	TTKEPAPTTKEPAPTTKKRAPPTPETPPTTSSEVSPTTKKEPT--TIHKSPOESP	854			
Db	1108	EYTVAPAEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAP	1166			
OY	855	ELSAEPYPKA-LENSKEPG-----VPPTKPPAAKPMETMTAKDKTTERDLRTTP----	904			
Db	1167	YAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAP	1226			
OY	905	EYTTAAPAKMETATTTEKTSKITANTTOYSTTODTT-----PKIIT--	951			
Db	1227	EETIYAP-----TEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAPTEETIYAP	1281			
OY	952	LKTTTLAPKYTTTKKIITTELMNNPDEETAKKDRAAT--NSKATTPKP-QKPTAPKPKPT	1008			

```

Ddb      1282 TEETTYAPTEATYTAPEETPYAETEETTEEPPGETTYYATEETTYTAPEETTYANMEET 1341
Qy       1009 -----SKRK-KTMPRVKREKTPTPRKMTSTWELNPTSRIAEMALOTTTPR 1055
Db        1342 PYEAEESTSYVSEKECNHEEFDEPDETDPDE--PSDEPTDEPDDEPTDLTDPDESP 1399
Qy        1056 --NOTPNKLEVNPKSDEDAG 1074
Db         1400 CDNOCINGIGVENKRYNNMAG 1420

RESULT   9
ID Q20007 PRELIMINARY; PRT: 1274 AA.
AC Q20007;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE COSMID F35A5.
DS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitia; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2; PubMed=7906398;
RX MEDLINE=94150718; Pubmed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Cratton M., Dear S., Du Z., Durbin R., Favallo A., Fulton L.,
RA Gardner A., Green P., Hawkins J., Kirsten J., Laister P.,
RA Jones M., Kerhaw J., Kirsten J., Mortimore B., O'Callaghan M.,
RA Lightning J., Lloyd C., Mcmurray A., Moore A., Saunders D., Showkeen R.,
RA Parsons J., Percy C., Rifken L., Roopra A., Smith A., Sulston J.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston R.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Leimbach D.;
RN Submitted (JAN-1996) to the EMBL/genbank/DDBJ databases.
RL [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RN Submitted (JAN-1996) to the EMBL/genbank/DDBJ databases.
RL [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RN Submitted (APR-1997) to the EMBL/genbank/DDBJ databases.
DR EMBL: U46675; AAB52641.1; -.
SQ SEQUENCE 1274 AA; 138065 MW; 07FB0DD292C1799F CRC64;

Query Match          11.4%; Score 846; DB 5; Length 1274;
Best Local Similarity 27.6%; Pred. No. 1,le-46;
Matches 313; Conservative 119; Mismatches 455; Indels 248; Gaps 60;

Qy     90 PSSKKAPPSSASQTISKTKRSP---KPP---NKKTKKVLESEITTEHSHSENSOES 142
    ||||| | : : : : ||| : : : : : : : : : : : : : : : : : : : : :
Db    185 PSPKKAAPSKEHDPIVPPTPIKNPAKKMKPPMEDDEVPTSELIEKPBPATKKVALKKKEP 244
    ||||| | : : : : ||| : : : : : : : : : : : : : : : : : : : : :
    143 SSSSSSSSSSTTIKKIKSSKSANRELQKLKYDKN-KNRTRK-KP-----TP----- 190
    SSSSSSSSSSTTIKKIKSSKSANRELQKLKYDKN-KNRTRK-KP-----TP-----
Db    245 STSVKPVSDSEPTKYV-----PVKLEPEVPPTPIKNPTKMRKPPWEDETPVEEVK 294
```

```
QY 191 KPPVNDAGSGDNGD----FKVTPDSTTGHNVSTSPKITTAKPI-NPRSLPNSD 245
D 295 EPEVPEKKAPVLLKKDPAPAAKADSPSKAAKPESSPVVPPVPPVKNPKKKKPPWE 354
QY 246 TSKETSLTVNKETVETKETTTNNKOTSTDGKEKTTSAKETOSIEKTSKADLAPTSUYLA 305
D 355 VDDEPAEYKKOSAPKKTPVLAKRKPEPSSSTPPSSDPSPKAAAVKRRDSSPKKAPPL 414
QY 306 KPPKAE-----TTTTGPA-----LTPKEKPTPTP-----KEPASTPK 340
D 415 QADPKAOEVPVPVKNPKVKKKPPWEVDEDEVEVEVKQDEAPAKKTPVLAKRKEPAKDTA 474
QY 341 EP-----TP-TTAKSAPTTKEPAPTTTSAPPTKEPAPTTKEPA----- 381
D 475 KPAKSTPTPEPKDPVKKPRDSSPKKVAAPKPDQAAP-TPVKNPVKKMRPWEDETERA 533
QY 382 -----PTTKEPAPTTKEPAPT-----TTKSAPTTKPEP-----APTTPKAPPTP 424
D 534 DDVSKPTDOKKTTPLAKKDPAPAKESLKPADTKAPAKPRDPSKKVAPTAPEKKTPLYA 593
QY 425 KE-----PAPTPKEPTPTTPKKEPAPTTKEPAPTTKEP-EPAPTAPEKAPAP-----TTPE 473
D 594 KKEPAGPADSKTKEPEKSKPRDPSKKAAPVAKPVKTEVAAVAKKPEPISKPDATAPK 653
QY 474 PAPTPKEPAPTTKEP-----SPTTPKEPAPTTTSAPTTKEPAPTTTSAPPTP-KEP 528
D 654 AEPNSPVVP-EPVKNPVKKKPPWEEDDAPAKPVSLPEPEKK-TPLYLAKKAPTKPDSEA 711
QY 559 SPTTTPKEPAPTTKEP-EPAPTTTPKAPAP-TPPKEPAPTTKEPAPTTKEPAPTA-PKE 583
D 712 AADPVSGSSKDPKLAKKAPKPDSPKMAVPIKPAKT--EVPVAVKKPEPVAKSRD 769
QY 584 PAPTPKEPAPTTKEPKLPTTP-----EKLAPTTPEKAPPTTPELAPTTPEEP-----T 633
D 770 PSPKAK-AEENSP--VVPPTPVKNPVKKKPPWEEDDAPAEVAVNPEPEKKTPLYLAKKI 826
QY 634 PTTPEEPAPTTPKAAAPPTTKEPAPTTKEPAPTT-KEPAPTTKEPAPTTKEPAPTT 692
D 827 PVKPRDPSKKAAPVAKKSTKTDAPVSKKPEVSKKPESSPKAEPNSPVVP---PTP 882
QY 693 LKEPAPTTTPKAPAKELAAPT-TTKEPTSTSDKAPPTTTPGTAPTTTPKEPAPTTKEPA 751
D 883 VKNVKKM-KPPWEDEDEPTEEVKKPSE--PEKKTPLYLAK-KEPEKKD-APVAVAKPRD 937
QY 752 TTPGTAPTTTPKEPAPT-----TPKKAPAP-----KELAPTTTKEPTS-----TSDK 793
D 938 PSPKAPPE--KEPAKAAKAPRDLSPKKAIPIPANTOEADPTPVKNPVKKKPPWEEDDE 995
QY 794 P-----APTTPKET-----APTTPKEPAPTTTPKEPAPTTTPPTTSEVSTPTTKEP 842
D 996 PAEVSAPAEDEKKTPLYLAKKAPKPRD--SPKKAAPVAK-PDKKIEPV-PPVTVKNP 1050
QY 843 TTIHKSP-----DESTEPISA-EP--TPKALENSPEKGV-----PTTKTPA 881
D 1051 VKMKKPPWEEDDSEVSAPEPEKKTPLYLAKKAPKPRATKPDSEAAADPVSGSTKDPK 1110
QY 882 AT-----KPEMTTAKDKTTERDLRTPTPT--TAAPMKTEATATTEKTESKTTAT 933
D 1111 LSKKAPVEKKTPTDOKLKSAPAKKPEKAPAEPAKMKKFWDDDDPEPADPTVPA 1170
QY 934 TQVSTTTTODTTFKITTLLKTTTTLADKVTTKTTTITTEIMNKPBEETAKPRATNSKAT 993
D 1171 PSKKPTEPADPLG-----GPKTKDPK-----LKKKAPAEKPTK----- 1206
QY 994 TPKPQKPTAPKAPKPTSTKPKTPMYRKPK-----TTPPKKMTSTMP 1036
D 1207 -PKKXVSKNEPKPTPEPKP-AAFKKMKPPWEEDDDEPADFTMPAKKPDTEP 1259
RESULT 10
QSPMO
ID QSPMO PRELIMINARY: PRT: 1315 AA.
```

```
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE EXTENSIN-LIKE PROTEIN.
GN PEX2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=POLLEN;
RA Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA Horst D., Gao M., Showalter A., Bedinger P.A.;
RT "Pollen Extensin-like (Pex) Genes in a Monocot and a Dicot.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF159297; AAD55980.1; -.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR001988; Xylose_Isom.
DR InterPro: IPR002965; P-rich_extenso.
DR InterPro: IPR003592; LRR_out.
DR Pfam: PF00560; LRR_3.
DR PRINTS: PR01217; PRICHEXTENSIN.
DR PROSITE: PS00172; XYLOSE-ISOMERASE-1; UNKNOWN_1.
DR SMART: SM00370; LRR; 3.
SQ
SEQUENCE 1315 AA: 134401 MW: 64C97A2A01F0936F CRC64;

Query Match 11.0%; Score 816; DB 10; Length 1315;
Best local similarity 23.5%; Pred. No. 9,9e-45;
Matches 265; Conservative 106; Mismatches 486; Indels 270; Gaps 35;

QY 49 RCSESEKRECCDOAOCKKYDCPCDYSCAEVHNPTSPSSK-KAPPSGASOTIKS 107
D 390 QCAPVLARVEVDCS-----KHVCAGPTPGCGRSSSVGKRPVSK----- 432
QY 108 TTKRSKPPNNKKTKKVISEITEHSYENQESSSSSSSSSTTWIKISSNKAAN 167
D 433 --PAAPAPMTHTPPDVSPEPLPESPV-----PAPAPMPTTLASPPAD 476
QY 168 RELQKKLVKDKNKKRKKPPKPPVNDAGSGDNGDFKVTPTDSTGHNVSTSPK 227
D 477 EYI-----PTPPVPAKSPG-----TSPASRGA-PLQADPP 508
QY 228 ITTAKPINRSLPNSDTSKETSITVNEKETTETTTNNKOTSTDGKEKTTSAKETQ 287
D 509 AASSPPAPPVSSPPPAV----- 527
QY 288 SIEKTSANDLAFTSUYLAKPTPKAETTTKGPALITTPKEPTPTTKEPASTTPKEPTPT 347
D 528 -VLPPAKTTPSPPAV-ASPPPEAVSSPOOVASPPPAVAVASPPPMKSP--PPAPV 583
QY 348 KSAPTTKEPAPTTTSAPTTT-KEPA-----TTTKEPAPTTKEPAPTTTPKEPA 399
D 584 ASPPPLMKSPPPPAVAVASPPQPLKSPPPVLMSTPSVKSPPPPVAVASPPPPVKSPPPL 643
QY 400 TTKSAPTTPE-----PAP--TTP-----KKRAPTTKEPAPTTKEP-PTPTTKEPA 448
D 644 APVSSPPPVLPPLPAAGKSTPPEEKPTPPPVKSSPPEKSLPPPLTSSPQEK 703
QY 449 EPAPTTKEPAPTTTPKAPPTTKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 508
D 704 PTPSTTSKPPPPSVENLPPPSKSSPPEEVSSPPQAKSSSP--PAPVS--SPP 756
QY 509 TTKAPPTTTSAPTTTKEPSPTTTPKAPPTTTPKEPAPTTTPKAPPTTTPKEP 568
D 757 LKSSPPPVSSPPPTKSSPPLAVSSPPQVEKTSPPPAVSSPPPTKSSPPLAVSS 816
QY 569 APPTTTPKAPTAPEPAPTTTPKEPAPTTTPKLPPTTPEKLAAPTTPKEPAPTTPELA 628
D 817 PPQVEKTSPPPAVSSPPPTPKSSPPLAVSSPPQVEKTSPPPAVSSPPPAVSSPPEK 876
```

QY	629	PEEPTPTPEEPAPTTKKAAPNTKEKAPATTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEGCT	668
Db	877	SVSSPPTTVKSSPPAPPLSSPPTPKSSSPPAHVSSPEAKSSSPPLAPLSSPSEKSP	936
QY	689	A-----PPTLKEPAPTTPKKPAKELAPTTTKEPSTSDKAPAT-----TPKG	732
Db	937	SSPMPVKTSPPAPVSSPPPTPKSSPP-----APVSSPPVVKSSPPAPVSSPPPTPKP	993
QY	733	TAPPTP-----KEPAPTTKEPAPTTPK-----GTAPTTLKEPAPTTPKPAKELAPTT	782
Db	994	LPPAPVSSPPPVKSSPPTPVSSPPTPKPLPPEPTPLVSSPPTPKKLPP-----APVS	1050
QY	783	TKGPTSTSDKAPATT--PKETAPTPK-----EPAPTTPKKAPATPTPTPTTSEVSTP	836
Db	1051	SPPPVVKSSPPAPVSLPPTPKSPPKRVSSPPRVKCCPPPLVSSPPAPKSLPP	1110
QY	837	TTTKEPTTIHKSPESTPELSAPTPKALENSEKEGVTPTTPATAPREMTTAKDKTT	896
Db	1111	TPVSSP-----PPEVKSSPPTPVSSPPAPKSSPPTPVSSPP-----	1150
QY	897	ERDLTTP-----ETTAPAKMTKETATTTTEKTESKITATTTQVTSITTDTPFKITT	951
Db	1151	--ELKSSPPAPVSSPPSAKSSPPAPVSLPPEKSSPPAPLSSPPEAKSP-----	1203
QY	952	LKTTTLAKVTTTKTITTTTEINMKKPEETAKPKDRATNSKATTPKPKAKKPTSTK	1011
Db	1204	-----PPPA-----PMSSLPPEVKSSPPAPVSSPPPKMSSPPPPAPIS	1242
QY	1012	KEKTPAPRVKPKPTTPPKMTSTPMLNPTSTRIAEAMLQTTTRPNOTPNKSLVEVNPSE	1071
Db	1243	SPPPAP--VKPSPSLPPAPVSSPPAPV-----TSAP-----PKKE	1275
QY	1072	DAGGABGEPHMLLRPHVMEVTPDDMDYLPRVNOGIIINPLASDE	1118
Db	1276	EDSTA-----PPADALPPSEFNDIILPIMANK	1303
RESULT 11			
ID	P70670	PRELIMINARY:	PRT: 2187 AA.
AC	P70670;		
DT	01-FEB-1997 (TREMBLrel. 02. Created)		
DT	01-FEB-1997 (TREMBLrel. 02. Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17. Last annotation update)		
DE	NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,		
DE	MUSCLE-SPECIFIC FORM GP220).		
OS	NACA.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=66312450. PubMed=8698236;		
RA	Yocov W.V., St-Arnaud R.,		
RT	Differential splicing in of a proline-rich exon converts alphaNAC		
RT	into a muscle-specific transcription factor."		
RL	Genes Dev. 10:1763-1772(1996).		
DR	EMBL: U48364; AAB18734.1; -		
DR	EMBL: U48363; AAB18732.1; -		
DR	MED: MGI:106095; Naca.		
DR	InterPro: IPR002715; NAC.		
DR	InterPro: IPR003037; TS-N.		
DR	Pfam: PF01849; NAC; 1.		
DR	Pfam: PF02094; TS-N; 1.		
SO	SEQUENCE 2187 AA; 220599 MW; 003646A864DEBFD CRC64;		

Query Match	11.0%;	Score 812;	DB 11;	Length 2187;
Best Local Similarity	27.4%;	Pred. No. 3e-44;		
Matches 311;	Conservative 128;	Mismatches 449;	Indels 246;	Gaps

Db	746	SPF-PPSSGAVPNTGA-----PPSKGAPVPTPSSSSQVPAELLBPQ	792
Qy	145	SSSSSSSSSTIMIKISKNSANRELQKLVKDNKKNTKPKPPVVDAGSGIDN	204
Db	793	KTEBVYASRLISAVQSPKVDPLMSD-----VPTSPKTSATAPVK-----	833
Qy	205	GDRKVTTPTDSTTOHKN-----VTSISKITAKIINRSLPNDSOTSETSI--TVNKE	257
Db	834	-----DTSATLSKVPAVATSLSPRPKAVASNAATVTPLEPITSLKNLAALATRE	885
Qy	258	T-----TVTKETTTNNKQSTDKCEKRTTSAKETOSIEKTSAKDLAPTISKVLAPPKA	311
Db	886	TLATSLPKVYTSPPQCTPKSVSLKAPAMTSKKMT---ELIASQVSPSQ--EPKEVPLL	940
Qy	312	ETTTKGPAITTPKEPTPTTPKKEPASTTEKEPTPTTISKAPITPKE--PAPITTKSAPITPK	370
Db	941	QHV-----PPTSPPKSPVSDTLISGALTSPPPKGPAPITLAETPTPKSKSPKPAASKKTPPATPS	997
Qy	371	-----EPAPITTKKEPAPITTPKEPAPITTKKEPAPIT--TKSAPIT-----	406
Db	998	PEGVAVPVLIPPCSKKAPKAPTAAPRESSITSSKKAAPTAVSKRIPSGVTAVPLEISLP	1057
Qy	407	-----TKEPAPITPKK--PAPITPKKEPAPITKEPTPTTPKEPAPITTKKEPAPITPKE	457
Db	1058	LKETSKAOTGEKASASPKSPKSPKTAGPKE---TPGCGVTAVPLEISLPKETFOAMTPE	1114
Qy	458	PAPTAAPKKAPITP--TKEPAP-----TTPKEPAPITTKKEPSPITPK-----EPAPITTKSA	506
Db	1115	SLAASSOKRSKPTSVPEPTPPGGVTAAMPLEIPSADOKAPKTAIVKQIPTEDAVITLIAGS	1174
Qy	507	PTTKKEAPITPTT--KSAPITPK-----EPSPTTKKEPAPITTKKEPAPITPKKAPITP--	556
Db	1175	PLSPKKASKTAAPKEAPATISGVAVISELSPSPKKTSSKTAAPKENSATLPLPKRSFKTA	1234
Qy	557	-PKE-----PAPITTKKEPAP-----TTTKKPAPTAAPKEPAPITTKETAPITP	596
Db	1235	APKEPATPSSEGVAVPSELSPSPPPASKGVAVLITPAGANLAE--SPSPSKVYKTA	1293
Qy	597	PKKLITPTPEKLAPITPKEK--PAPITPEBELAPITPEEPTPTT-----PREPAPITPKA	647
Db	1294	ADEESTTP-----SQOKLIPKAVGREASAITPPSKIKPITAKIVREMSAPSEGVTAVPLE	1347
Qy	648	AAPNTPKEPAPITPKE--PAPITKEPAPITPKETAPTTPKGTAPITTKKEPAPITP-----	701
Db	1348	IPSPRKAPKTAAPKEPAPPS--PEGATTPAVQIPSPSKSGSKAGSKE--TPPTSPSEGV	1404
Qy	702	-----KKPAPRELAPTTKTPPS-----TSDKRAPITPKGTAPITPKE	740
Db	1405	TAAPLEIPISSKKTSKMASPRELVTPSSKLSQIVGKRETSLECATVAPLEIPPSHKA	1464
Qy	741	PAPITPKE--PAPITPKGTAPITTKKEPAPITPKKPAKRELAPTTKGTPTSTSDKAPITP	799
Db	1465	PKTVDPKQVPLIPSPK--DAPITLAE--SPSSPK--APKTAAPSESR--VITVPEPKKA--TPQ	1519
Qy	800	KETAPITTKEPAPITPKKAPATPTPEP-----PPTISEVSTPTPTK-----EPTTIKS	848
Db	1520	KASGTTASKVVPAPATQEVAAVSSRETVTPAVPVKPPSSHKKTSTTLEKEAPITLPPS	1579
Qy	849	PDGSTEPLSAEPTPALENSKE--PGVPTTKTAPAAKPEMTTAAKDKTTERDLRTPETTY	907
Db	1580	PKSPKSPISPKKAPIT--SAPKEPAPSPSIR--PVTT--SLAQAPPSLOKAPSTIIPKEN	1634
Qy	908	TAAPKMKETATTEKTESKITATTOVISTTQDTTPKIKITLTLTLAKVYTTTKT	967
Db	1635	LAPAV-----LPIVSSKSPAAAPARASASISLPTAAPQJAPKEATVITPSCKKAATETPIET	1690
Qy	968	ITTEILMNPPEEAPAKPDRATSKATTPKPKAPKAPKKTSTKCEKTPMRYKPKTPTT	1027
Db	1691	STAPSLGAKETSE-----TYSKILYMSPP-----PKASSSKRABSTP-----	1730
Qy	1028	PKMTSMELNPTSRILAEAMLQTTTRPNQTPNSKLVAVPKSEDDGAGGETP	1081

```
Db      1691 STAPLEGAPKETSE-----TSVSKYLWSSP-----PKKASSKKRASTLP-----1730
QY     1028 PRRKMTSMELNPTSRIAEAMLOTTTRNQTPNSKLVEVPKRSEDGAGAEETP 1081
```

Db 1731 ----ATTLESLKASVLS-----PTATSGKSHSPVS-DACSTGTTTP 1770

RESULT 12

09LIE8 PRELIMINARY; PRT; 1480 AA.

AC 09LIE8; 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DE 01-JUN-2001 (Tremblrel. 17, last annotation update)

OS SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.

OC Arabidopsis thaliana (mouse-ear cress).

CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA.

RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.; Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA.

RX PubMed=10907853;

RA Nakamura Y.

RT "Structural analysis of Arabidopsis thaliana chromosome 3. II. Sequence features of the regions of 4,251,695 bp covered by ninety P1, TAC and BAC clones."

RT DNA Res. 7:217-221(2000).

RL EMBL: AP001306; BAB03062.1; -

DR InterPro: IPR002965; P-rich_extensn.

DR PRINTS: PRO1217; FRICHEXTENS.

SO SEQUENCE 1480 AA; 147153 MW; DIACOC79F155E732 CRC64;

Query Match 11.0%; Score 811.5; DB 10; Length 1480;
Best local similarity 28.8%; Pred. No. 2.2e-44;
Matches 330; Conservative 60; Mismatches 539; Indels 215; Gaps 51;

QY 86 PTPSPSKAAPPSSGASQTIKSTTKR--SPKPPN-----KKTKKVIYSESEIEHS 135
Db 46 PPKPVPVKRPHNPK--PRTIKRRPRKRNHPRKPVYKRNHPRKPRKRNHPRKPPK- 102

QY 136 VSENOSSSSSSSSSSSTIMKIKSSKSNANRELOKILVK--DNKKRITKKKPKPP 193
Db 103 -----PHNKPRTIKPPRPHK---PRPHKPPNVKRNHPRKPPRTKRNHPRKPP 146

QY 194 VDEAGSGLDNDGFKVTTEDTSTQHNKYSTSPKITTAKPINRPSLPSNDSKETSILT 253
Db 147 -----TKNH---PHKPPRTIKPPKPPSVKPPSPTPK----- 175

QY 254 VNKETTVETKETTITTKOTSTGKEKTSIAKETOSEKTSANLADPTSKV--IAKPPTKA 311
Db 176 -----PPTNPNPSTQPRPHKPPRCPTRPRVPSPMANRPMQPIANR-PIA 223

QY 312 ETTTKGPAITTPKEPPT--TTP--KEPASTPKEPPTTIKAKPTTPKEPAPT-- 362
Db 224 KSPVATPPTATPPTATPPTITPPTVATPPTTPIINPPIIMPPIATPVAAPRIPNIS 283

QY 363 KSAPTTPKEPAPT--KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 401
Db 284 KPPVTPP-----PTTTPPIAKPPIATPPISTPRAATPRAATPPTITLPPAKPPVATSPVIT 339

QY 402 KSAPTTP--KEPAPTTPKAPAP--TPKKEPAPTTPKEPPTTPKEPAPTTPKEPAPTTP-KE 457
Db 340 --PPTPTIAQPPVATPPTATPPTVATPPTATPPTSKSPSTPISSEPVATPPTATSPIKT 397

QY 458 PAPTPKPPAPTTPKEPAPTTPKEPAPTTP--KEPAPTTPKEPAPTTPSAPTTPKEP 513
Db 398 PPKAPPVATPPIAKSPIATPPTATPVAATPPIEIKPPVATPPTTTPPIAT--PVAKPP 454

QY 514 APPTTKSAPTTPKPP--SPITTKKEPAPTTPKEPAP--TTPKKAPAPTTPKEPAPTTPKEP 568

Db 455 VETPPIATPPTAKPPISTPPISSKPPVATPRAANPPIITTPPRVAPPTATPPIAPKPP 514
QY 569 --APPTTKKRAPAP--KEPAPTTPKEPAPT--PKLTP--TPEKLAPTTPKEP 616
Db 515 VTPPTATPPIATPPIAKSSVPATPPTATPPTVATPPIAKPPVVPPTTPTTATPVAKPP 574

QY 617 APPTPELAPT--PEEPTPTTPEEPAPTTPKAAPNT--PKPAPTTPKEPAPTTPKEP 672
Db 575 VAPPIATPPTAKPPISTPPISSVPATPRAAPPTTTPPPAKPPVATPPIATPPIAKPP 634

QY 673 APPTPELAPT--TTPGTAAPT--LKEPAPTTPKPAKELAPTTPKEPPTTSDDKAP 727
Db 635 VAPPIATPPIATPPIAKPPVATPPIKTTPPAKPPVAIPPIATPVAKPPVAT--PPTA 691

QY 728 TTPKGTAP--TTPKEPAPTTPKEPAPTTPGTAAPTTLKEPAPTTPKPAKELAPTTPK 786
Db 692 TTPPIATPPIATPPTVTPPTATPPTVATPPIAKPPPTI--PPTATPVAAMPPIATPPIAKPP 749

QY 787 TST--TSDKPAPTTPKEPAPT--TPKKEPAPTTPKPAPTTP-----ETPPTTSEVSTPT 837
Db 750 IATPPIATPVAKPPVATPPTATPPIATPPIAKSPVATPPTATPPIAKPPVATTP 809

QY 838 TTPKEPTTIH--KSPDESTBELSAEPKPKLENSPKKGVPTTKTPAATKPEMT--TTAKDK 894
Db 810 TAPPTATPPIAKPPVATPPIATPPIAKPPILTPPISKPPVATPRAANPPIITTPPPAKPP 869

QY 895 TTERDLRTTPETTTAARPKMTKETATTEKTESKITAATTQVTSSTTQDDTPPKITTLTKT 954
Db 870 VATPPIATPP--IAKPPVATPPTATPPIATSPVAKPPVAIPPIKTPPKPPVATPIAT 927

QY 955 TILAKVTTTKT-----ITTEIMKPEETAKPKORATNSKATTPPKPPKAPKPTST 1010
Db 928 PPAKPPVATPPTATPPIATSPDIATPPIVAT--PPTATSPATPPIAKPPPTTP--PPT 982

QY 1011 KKPATPVRKPKETTPPTPKMTSTMPELN-----TSRAE-----AMQITTR 1054
Db 983 -PPVAMPPIAP--PTAKPPVATPPIANPVEKPPVATPPIAKPPVLPPIAKPPVETS 1038

QY 1055 PNOTNSKL--VEVVPKSEDAGAGETPRHMLLRPHVEMPEVTPDMOVLPRVNOGIIIN 1112
Db 1039 PPAATPPIATPPIAP--VVKPPVAIPPTIPPAVATPPTVTPPRAAMP 1083

QY 1113 PMLS 1116
Db 1084 PIVT 1087

RESULT 13

041805 PRELIMINARY; PRT; 1188 AA.

AC 041805; 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DE 01-JUN-2001 (Tremblrel. 17, last annotation update)

OS EXTENSIN-LIKE PROTEIN PRECURSOR.

OC Zea mays (Maize).

CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; OC Penicillaceae; Andropogoneae; Zea.

OX NCBI_TaxID=4577;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=873; TISSUE=POLLEN.

RA Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.; Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.

RL EMBL: Z34465; CAA84230.1; -

DR Mendel; 14346; Zeama;2368;14346.

DR InterPro: IPR001611; LRR.

DR Pfam: PF003592; LRR_out.

DR SMART; SM00370; LRR; 4. signal.

FT SIGNAL 1 27 POTENTIAL..
 SQ SEQUENCE 1188 AA; 120980 MM; 2C77C7F8D7130149 CRC64;
 Query Match 10.6%; Score 784; DB 10; Length 1188;
 Best Local Similarity 29.9%; Pred. No. 1e-42;
 Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;
 188 PPPKPVNDAGSGLDNGEFKTTPTSTTQHNKYSTKITTAKINRPLST---PPN 243
 460 PPPHSPDP-----DYPPPTPEVPEKSPAPNSPSQVQPPAPASTPPPLVNLSPQQ 510
 244 SDTSKETSILVNKETTVEKTTTNTKQSTDKKETSAKETQSTLEKTSANDLAPTSKV 303
 511 APVG-----SPPPVKTSPAPAPIG---SPSPPPVSVYV 541
 304 ----LAKPPKAETTKGAPLTTKEPPTTTPKEAPSTTKEPPTTITKSAPTTKEPA 358
 542 SPPPVKSPPPAPVGSPPPEKSPPPAPVASSPPPVSP--PPPTLVASPPPVKSPSP 599
 359 PTTTKSAPTP--KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 415
 600 PPPAPVASSPPPVKSPPPPTPVASPPPPAPVASSPPPMKSPPTTTPVSSPPPEKSPPPP 659
 416 PKKPAPTTKEPAPTT-----PKE---PPTTTPKEPAPTTTKEPAPTTTKEPAPTTKPKP 466
 660 PAKSTPEPEYPTPTPEVSKSSPPEKSLPPTLIPSPPOEKPPTPSTPSKP--PSSPEKP 718
 467 APPTTKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 526
 719 SP--PKPEVSSPQTP---KSSPPAPVSSPPTTPVSSPPALAPVSSSPVASSP---- 768
 527 EPSPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 586
 769 PPAPLSSPPAPVQVKSPPPVQVSSP--PPAKSSPPAPVSSPPOYEKTSPPAPLSSSP 827
 587 TTPKETAP---TTP---KKLTPTPEKLAPTTPEKAP-----TTPPELATTPPEPT 633
 828 LAPKSSPPHHVSSPPPVVSSPPAPVSSPPLTPKASPPAHVSSPPEVVKSTP--PA 885
 634 PTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 692
 886 PTVYISP--PSEPKSSPPTPVSLPPPIVKSPPAPVSSP--PMMPKSSPPPVVSSPPTT 943
 693 LKE---PA-----PTTPEK---PAKELAPTTTKE---PTTSTSDKAPPTTKEKTAATTP 738
 944 VKSSPPAPVSSPPATPKSSPPAPVNLPEPEVSSPPTTIVSSPPA---PKSSPPAP 1000
 739 -KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 797
 1001 MSSPPPVKSPPPAPVSSPPPVKSPPPAPVSSP---PPVKSPPPPAPVSSPPPV 1057
 798 -TPKETATTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 856
 1058 KSPPPAPLSSPPPVKSPPPAPVSS--SPPPVKSPPPAPVSSPPPTTIVSSPPAPVSS 1115
 857 SAEPPPKALENSPKPEGVPTTKTPATKPEMTTAKDKTTERDLRTTPEPT 906
 1116 SPPAP-----VKPPSLP--PPAPVSSPPPVVTPAPPKKEOSLPPAPES 1188
 RESULT 14
 09VEL9 PRELIMINARY: PRT: 2112 AA.
 AC 09VEL9
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CG4090 PROTEIN.
 GN CG4090
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Andrews-Pfannkoch C., Baldwin D.,
 RA Abell J.F., Agbayani A., An H.-J., Bayraktaroglu L., Beasley E.M.,
 RA Ballew R.M., Basu A., Baxendale J., Bhardwaj D., Bolshakov S.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Brokstein P., Brotier P.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Butlis K.C., Busam D.A., Butler H., Cadenot L.B., Davies P.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dudin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko F., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon R., Nuskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relvert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier A.R., Spirdling A.C., Stapleton M., Strong R., Sun E.,
 RA Switskas R., Tector C., Turner R., Ventier E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003716; AAF55402.1;
 DR FlyBase: FBgn0038492; CG4090.
 DR InterPro: IPR002557; Chitin_binding.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF01607; Chitin_bind_2; 10.
 DR SMART: SM00494; Chitin2; 11.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS00022; EGF_1; 1.
 SQ SEQUENCE 2112 AA; B91018E551A5D36 CRC64;
 Query Match 10.6%; Score 782.5; DB 5; Length 2112;
 Best Local Similarity 27.1%; Pred. No. 2.4e-42;
 Matches 310; Conservative 106; Mismatches 369; Indels 361; Gaps 49;
 54 FERGRCDCD-----AOCKKYDKCCPDYESCAEYHNFTSPSSKKRAPPSSGASQITK 106
 969 FKPAEKCESETEFLADNENCSKEYRCVNDGNGKGFTRY-SFTCPNTLMDPEANSHHPDO 1027
 107 STTKRSKPPNKKTKKVI-----ESEETHEHVSENOESSSSSSSSSTITWKIKSSK 162
 1028 IOTK-----PLCKKRVVSGGSSSNSTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 1074
 163 NSANRELQKKLVKDKNKKRKKKPPKPPVVDAGSGLDNGDFKVTYTPDITSTQHNKV 222
 1075 NS-----GSSSNTGSSSSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSS 1100
 223 STSKRTTAKINRPLSPNSDTSKETSILVNKETTVEKTTTNTKQSTDKKETS 282
 1101 SSS-----NSGSSSGSSSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSS 1138


```

QY 263 AKETOSIETSAK-DLAPTSKVLAKPTKAEPTTK-----GP----- 318
Db 1139 SSSSSSSSTSKKPNPSETCKVNGQFIDGRSDCAFYRCVNDRCGMNPFSCGPGVW 1138
QY 319 -----ALTPKPEPTPTTPEKPASTTKKEPTPTTISKAPPTTKEPAPTT 361
Db 1199 DAQMOMAHNAVKECGGIAPPTTSTPTTSR-PTTASTSRSDQTSRPRGP-----PTT 1253
QY 362 TKSAPPTTKEPAPTTPTTKEPAPTTKEPAPTT----- 392
Db 1254 ARPVATPPTSSPTTASSQTTSPYQAPNDGCRSGFADPNNSKFRVCRNNKG 1313
QY 393 -TKPE-----APT-----TKSAPPTTKEPAPTTPTTKEPAPTT 431
Db 1314 FTSTPFCGAGTVDODLOTGNCNHNFNNSGTGTESTTTPKPC--EPATNGTATSTSTTT 1371
QY 432 PKPEPTPTTKEPAPTTPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPS 491
Db 1372 P--PPTTDLPTSTTGLP-PTTTELPLP-----TTTDLPTTTLRLPPTTTLPLP 1421
QY 492 PTPKPEAPPTTTSKAPPTTKEPAPTTTSKAPPTTK--EPSPPTTKEPAP----- 538
Db 1422 TTTTGLPPTTTGAQPTTTTLSTETSTVTTSPESTTOPPSTTTMKPLAGTCTGEGY 1481
QY 539 -----TPK-----EPAP-----TPKKAAPT 556
Db 1482 MADEDCRKYKRCINAGASYRKYNFTCPKGTGWNVEVQTCDYVENIIPRCSKLPAEPTTT 1541
QY 557 P----KEPAPTTKE--PAPTPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKE 611
Db 1542 PSEESKOPGSTTQSTDEPTTVPKPI-TKPE--EPSTEKQOKPPTOYPEKPTTEE--PE 1597
QY 612 TPKEPAPTTKEELAPPTTKEPTTPEEPAPPTTKEAAPTPTTKEPAPTTKEPAPTT-PK 670
Db 1598 KPAKPT-TPPEYPOKPTTPEEPPEKPOKPT-TPPEYPOKPTTPEEPPEKPOKPTTPEYPO 1655
QY 671 EPAPTTPEKTA-----PTPK--GTAPTTKEPAPTTTPKPAKELAPPTTKEPTSTSD 723
Db 1656 KPTTEPTTTISIGYNPTTTSVPGYNPTTTPRIVETTTSTPGYK--PTTGTBPTTTT- 1711
QY 724 KPAPTTTPKGAAPTPTTKEPAPTTKE--PAPTTPKGAAPTTLKEPAPTTTPK--PAPKE 777
Db 1712 LPSTTTDAOEPTTTSKPEPTTTTESPESTTSGSVTTIQPEPQMYNCSSECFEPD 1771
QY 778 LA-----PTTKGPTSTSDKAPPTTPEKTAAPTPTTKEPAPTTTPKPAPTTPEPTPTS 831
Db 1772 CSNYTCVDAKAKGYAYAFKCGKGTWDTSTETCNVADQVSGN-----CSSG 1820
QY 832 EVSTPTTKEPTTIHKSPESTPELSAAPTPKALENSPKPEGVPTTKTPAATKPEMTTA 891
Db 1821 QTTPTGTTTEPGTTESTSSGK-----ETTSKAPEN-----TTTWA--PETTTT- 1863
QY 892 KOKTTROLTPTTETTTAAPTKEATTTETTESKITATTTQVTTSTTODTTFKITTT 951
Db 1864 -----SSPETTTT--VASETTTTSSTG-----TTTAPETTTTPKPKP----- 1898
QY 952 LKTTTLAPVTTTKTITTTTEIMNKPDEETAKPKDRAATKATTPKQKPTKAPKPTSTK 1011
Db 1899 -ETTITAGETSTSKSPTTTE-----SPAPSTNTSAP----- 1929
QY 1012 KKKTAP 1017
Db 1930 CPEGTG 1935

```

RESULT 15

Q9XND2 PRELIMINARY: PRT: 763 AA.

Q9XND2: 01-NOV-1999 (TREMblrel. 12, Created)

DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)

DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)

```

DE PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_taxid=1773;
RN [1]
RC SEQUENCE FROM N.A.
RP SRRAIN-H37RV;
RA Esplia C., Lacleite J.L., Mondragon-Palomino M., Martens A.,
RA Zhang Y., Moreno C., Singh M.;
RT Cloning and characterization of a new member of the PGCS family that
RT is a useful marker of polymorphism in Mycobacterium tuberculosis.
RL Submitted (JUN-1998) to the EMBL/GenBank/DBD databases.
DR EMBL: AF071081; AFD41594.1;
DR InterPro: IPR002951; Atrophin.
DR InterPro: IPR002965; P-rich_extensin.
DR InterPro: IPR003882; Pistill_extensin.
DR PRINTS: PRO1222; ATROPHIN.
DR PRINTS: PRO1217; PRICHEXTENSIN.
DR PRINTS: PRO1218; PSTIEXTENSIN.
SQ SEQUENCE 763 AA; 75034 MW; 39168EC45A5916F8 CRC64;

```

Query Match 9.6%; Score 715; DB 2; Length 763;

Best Local Similarity 28.2%; Pred No. 1.9e-38;

Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

```

QY 307 PTPKAEPTTKGAPALTPKPEPTTTPKEPASTTPEKPEPTTISKAPPTTKEPAPTTTSKAP 366
Db 3 PVP-----AFRALALPLPAPAPAPAEKSKPPPPAP-----PAPCWMILVSAAP 46
QY 367 TTPKEPAPTTKEPAPTTTPKEPAPTTTPKSAAPTTPKEPAPTTTPKKAAPTTPKE 426
Db 47 PCR--PAP-----PAPPKSKAPPPVVPAPAPAPAPLPL--PAP-----DEADRE 90
QY 427 PAPPTTKEPTPTT-----PKEPAPTTKEPAPTTPK--EPATAPK--KPAPTTKEPAPTTPK 480
Db 91 SRPALPCCPPPPVVIIDPEEPAPAPVPAPNSPPPPPPAPKVFAPAPV--PVPNSPP 148
QY 481 EP--APTTKESPTTKEPAPTTTPKSAPTTTPKEPAPTTTSAPPTTPKEPSPTTKEPAP 538
Db 149 FPPFPALINPAP-----PAPPLANSPLPLPAPPTPAGT--PAPAPVPVPAAPAKSKPA 201
QY 539 TTPKEPAPTTTPKPAPTTTPKEPAPTTTPKEPAP-----TTTKKAPAPAPAPPTTPKEAP 594
Db 202 SPPRPAP-----PAPATPMEPPPLPVPVPPDISKETPAPAPAPAPVAPVPIPVPLP 256
QY 595 TTPKALPTTPEKLA-----TPPEKAPPTPEELAPPTTPEEPPTTPEEPAPPT-- 644
Db 257 PVPNKIIPAP--APVAVAVAVIAPCPPLPLPNNHPAPAPAPAPVGVGLAPLPSHP 313
QY 645 --PKAAPNTPKPEPAP----- 658
Db 314 PAPAPAPVGPVLAFLPISGRPVSVKGSFTTSTFCCRCVSGEVLALGALNPSRSP 373
QY 659 --TTPKEPAPTTTPKEPAPTTKEA-----PTTPKGA--PTTLKEPAPTTP--KPAPE 708
Db 374 TTTTALPLAPLPLPLPLPINTAVPPIPLPVTLAPLPLAPLPLPLSPGVPPAP-- 431
QY 709 LAPTTKEPTSTSDKAPPTTPKGAAPTTPKEPAP-----PTTKEPAPTTPKGAAPTTL 762
Db 432 --PAPGKFWTTPPLAPAPPEPK--TVPLPBGPCSPSEKPNPAPPEPEPSSPALP 488
QY 763 KEPAPTTP--KKPAPELAPTTTPKGP--TSTSDKAPPTTPKE--APTTPKEPAPTTP 815
Db 489 APAPASMSAVRVPSPPIPPAPAPAPASMPALPAPAPSPAPATRLCPPLPSPAPVNSP 548
QY 816 KKAAPTTPETPTTSTSTPTTTPKPTTTPKESD--DESTELSAPTPKA--LENSP 869
Db 549 --PAPAPPTTPKPLS--ANPPCPVVPAPAPAPAPAPAPAPAPAPAPAPPPPPVANS 604
QY 870 KEGVPTTKTPAATKPEMTTAKDKTERDLRTTPETTTAAAPKTEATTTKTESK 929

```

```

Db      605  --PAPPAPAPPSALPFVNBPB-----PPTPAPK-----SRPAL 637
OY      930  TATTQVSTTTQDTTPEFKITTLKTTTLAPKYTTTKTITTEIMNKPEETAKPKDRATN 989
          |      :      |||
Db      638  PAAPPAAPPAPVVRATTP-----PPAAPAPAPN 665
OY      990  SKATTRPKPKP---TKAPKKPTSTKKPKUMPRVRRKPTTPPRKMTSTM---PELNP 1040
          |      |      |      |      |      |      |      |      |
Db      666  SMALLPPAPDDPIPLLATPPAPAPPLPMSPPAPPLPPAAPDDPPAPPLITINOPSPPLAP 725
OY      1041 TSRIAEAMLQTTTRPNQTPNSKLVENVNPKSEDAAGAGGET 1080
          |      |      |      |      |      |      |      |      |
Db      726  VPGAPLAPLPIINGRPVFAKKNSLI-----GSSSGDT 756

```

Search completed: April 26, 2002, 16:22:02
 Job time: 507 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:22:18 ; Search time 153.5 Seconds

(without alignments)
1013.901 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1263
Perfect score: 5698
Sequence: 1 VQDNKKNRKKKPKPPVY.....VAALSTAKYKNPESVFEK 1064

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phase:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP Vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5698	100.0	1404	4	Q92954
2	5669	99.8	1404	4	Q9BX49
3	2519	44.2	1054	11	Q9JMG9
4	1048.5	18.4	401	6	Q77765
5	967.5	17.0	1079	5	Q9W4S7
6	936.5	16.4	1049	5	Q917S1
7	865.5	15.2	1795	5	Q76894
8	853	15.0	1489	10	Q96449
9	808.5	14.1	1315	10	Q9SPM0
10	802	14.1	1274	5	Q20007
11	800.5	14.0	1480	10	Q9LIR8
12	784	13.8	1188	10	Q41805
13	776	13.6	2187	11	P70670
14	739	13.0	2112	5	Q9VEJ9
15	715	12.5	763	2	Q9XDH2
16	707.5	12.4	555	10	Q9FP06
17	671	11.8	1151	13	Q57580
18	667.5	11.7	4880	11	Q9JFT1
19	667.5	11.7	5085	11	Q9JXS6

20	663.5	11.6	2284	5	Q9VPG1	Q9VPG1 drosophila
21	660.5	11.6	6677	5	Q9NA35	Q9NA35 caenorhabdi
22	659.5	11.6	1229	5	Q94185	Q94185 caenorhabdi
23	653.5	11.5	1514	5	Q9GHW7	Q9GHW7 leishmania
24	647	11.4	4833	11	Q9QXY6	Q9QXY6 mus musculu
25	647	11.4	5038	11	Q9QIX7	Q9QIX7 mus musculu
26	643	11.3	2089	4	Q14676	Q14676 homo sapien
27	633	11.1	489	10	Q41707	Q41707 vigna ungu
28	632.5	11.1	7962	4	Q10465	Q10465 homo sapien
29	632	11.1	761	10	Q9Z010	Q10465 homo sapien
30	632	11.1	6632	5	Q17362	Q17362 aradidopsis
31	629	11.0	3507	5	Q23587	Q23587 caenorhabdi
32	628.5	11.0	1612	5	Q9YQ02	Q9YQ02 drosophila
33	626.5	11.0	990	13	Q91803	Q91803 xenopus lae
34	625	11.0	6642	5	Q01761	Q01761 caenorhabdi
35	624.5	11.0	2768	5	Q9VC00	Q9VC00 drosophila
36	622.5	10.9	839	2	Q9RX57	Q9RX57 delinococcus
37	622	10.9	3570	4	Q99552	Q99552 homo sapien
38	620.5	10.9	971	5	Q9XV84	Q9XV84 caenorhabdi
39	615	10.8	2244	5	Q9N3Y8	Q9N3Y8 caenorhabdi
40	607.5	10.7	801	5	Q23635	Q23635 caenorhabdi
41	607.5	10.7	924	12	Q99307	Q99307 epstein-bar
42	605	10.6	379	5	Q27929	Q27929 drosophila
43	600.5	10.5	1893	5	Q9NKC9	Q9NKC9 drosophila
44	598.5	10.5	409	10	Q9SBM1	Q9SBM1 volvox cart
45	577	10.1	956	10	Q9LJ64	Q9LJ64 aradidopsis

ALIGNMENTS

RESULT 1
ID Q92954 PRELIMINARY; PRT: 1404 AA.
AC Q92954
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MEGAKARYOCYTE STIMULATING FACTOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Perenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhatia S., Kriz R., Hewick R., Clark S.C.,
RL Blood 78:279-279(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA Fitzgerald M., Scaltireto J., Kelleher K., Preissner K., Kriz R.,
RA Jacobs K., Turner K.,
RL (in) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL Mosher D.F. (eds.):
RL Biology of vitronectins and their receptors., pp.45-52,
RL Elsevier Science Publishers B.V. (1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,
RA Perenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhatia S., Kriz R., Hewick R., Clark S.C.,
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN EMBL: U70136; AAB09089.1; .
RP InterPro: IPR000585; Hemopexin.
DR InterPro: IPR001212; Somatomedin_B.
DR Pfam: PF00045; Hemopexin_2.
DR Pfam: PF01033; Somatomedin_B_2.
DR PRINTS: PR00438; GFCYSKNOT.

DR PRINTS: PR00022; SOMATOMEDINB.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
SO SEQUENCE 1404 AA; 151090 MM; AABD7AD19B35FAF6 CRC64;

Query Match 100.0%; Score 5698; DB 4; Length 1404;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKNRTRKKRPKPPVVDAGSGLDNGDFKVTTPDTSTTQHNKVSPTSKITAKPIN 60
DB 200 VKDNKKNRTRKKRPKPPVVDAGSGLDNGDFKVTTPDTSTTQHNKVSPTSKITAKPIN 259
QY 61 PRPSLPNSDTSKESLYNKETVEKETTNNKOTSDGKEKTSKETSQSIKTSK 120
DB 260 PRPSLPNSDTSKESLYNKETVEKETTNNKOTSDGKEKTSKETSQSIKTSK 319
QY 121 DLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTKEPAATTTKSAPTTTSKAPTTK 180
DB 320 DLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTKEPAATTTKSAPTTTSKAPTTK 379
QY 181 EPAPTTTSKAPTTKPEAPTTTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 240
DB 380 EPAPTTTSKAPTTKPEAPTTTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 439
QY 241 PKKAPTTTKEAPATTKEAPTTTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 300
DB 440 PKKAPTTTKEAPATTKEAPTTTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 499
QY 301 PTTKEAPATTTKEAPSTTKEAPATTTSKAPTTTKEAPATTTSKAPTTTKEAPSTT 360
DB 500 PTTKEAPATTTKEAPSTTKEAPATTTSKAPTTTKEAPATTTSKAPTTTKEAPSTT 559
QY 361 PAPPTTKEAPATTTKKAPTTTKEAPATTTKKAPATTTKKAPATTTKKAPATTTKKAP 420
DB 560 PAPPTTKEAPATTTKKAPTTTKEAPATTTKKAPATTTKKAPATTTKKAPATTTKKAP 619
QY 421 TPKKLTPTTPEKLAPTTPEKAPATTPEELAPTTPEEPTPTTPEEAPTTTKEAAPT 480
DB 620 TPKKLTPTTPEKLAPTTPEKAPATTPEELAPTTPEEPTPTTPEEAPTTTKEAAPT 679
QY 481 PAPPTTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 540
DB 680 PAPPTTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 739
QY 541 EPTSTTSKAPATTPKGAPTTTKEAPATTPKGAPTTTKEAPATTPKGAPTTTKEAPATT 600
DB 740 EPTSTTSKAPATTPKGAPTTTKEAPATTPKGAPTTTKEAPATTPKGAPTTTKEAPATT 799
QY 601 KELAPTTTSGPTSTSDKPAPTTKEAPATTPEEPTPTTPEEAPTTTKEAAPT 660
DB 800 KELAPTTTSGPTSTSDKPAPTTKEAPATTPEEPTPTTPEEAPTTTKEAAPT 859
QY 661 PTTTKEAPATTKEAPSTTKEAPATTTSKAPTTTKEAPATTTSKAPTTTKEAPSTT 720
DB 860 PTTTKEAPATTKEAPSTTKEAPATTTSKAPTTTKEAPATTTSKAPTTTKEAPSTT 919
QY 721 TERDRLTTPETTTAAPTKEATTTTKEKTSKATATTTOYSTTQDTTODTTPPKITL 780
DB 920 TERDRLTTPETTTAAPTKEATTTTKEKTSKATATTTOYSTTQDTTODTTPPKITL 979
QY 781 TLAPKVTYTTKTTTITTEINNKPEETAKPKDRATNSKATTPKOKRTKPKPTSTKKKT 840
DB 980 TLAPKVTYTTKTTTITTEINNKPEETAKPKDRATNSKATTPKOKRTKPKPTSTKKKT 1039
QY 841 MRRVRRPKTTPPKKMTSTMPKELNPTSRIAEAMLTTPKPNQPSKLVENPKSEDAG 900
DB 1040 MRRVRRPKTTPPKKMTSTMPKELNPTSRIAEAMLTTPKPNQPSKLVENPKSEDAG 1099
QY 901 AEGETPHMLLRPHVMEVTPMDADYLRPVNOCIIINPMLSDETNINCGKRPVGLTTLRN 960

DB 1100 AEGETPHMLLRPHVMEVTPMDADYLRPVNOCIIINPMLSDETNINCGKRPVGLTTLRN 1159
QY 961 GTLVAFRGHYFWMLSPSPSPSPARRITEWGIPSPIDVTTRONCGKTEFFEDSOQWRF 1020
DB 1160 GTLVAFRGHYFWMLSPSPSPSPARRITEWGIPSPIDVTTRONCGKTEFFEDSOQWRF 1219
QY 1021 TNDIKDAGYPRKIFKGGGLGQIVAAALSTAKKKNMPESEYVFK 1064
DB 1220 TNDIKDAGYPRKIFKGGGLGQIVAAALSTAKKKNMPESEYVFK 1263

RESULT 2
Q9BX49 PRELIMINARY; PRT; 1404 AA.
ID Q9BX49;
AC Q9BX49;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR).
GN BG174L6.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wray P.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL13553; CAC36090.1; --
SO SEQUENCE 1404 AA; 151076 MM; 782A11746B3FDEE5 CRC64;

Query Match 99.8%; Score 5689; DB 4; Length 1404;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1062; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VKDNKKNRTRKKRPKPPVVDAGSGLDNGDFKVTTPDTSTTQHNKVSPTSKITAKPIN 60
DB 200 VKDNKKNRTRKKRPKPPVVDAGSGLDNGDFKVTTPDTSTTQHNKVSPTSKITAKPIN 259
QY 61 PRPSLPNSDTSKESLYNKETVEKETTNNKOTSDGKEKTSKETSQSIKTSK 120
DB 260 PRPSLPNSDTSKESLYNKETVEKETTNNKOTSDGKEKTSKETSQSIKTSK 319
QY 121 DLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTKEPAATTTKSAPTTTSKAPTTK 180
DB 320 DLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTKEPAATTTKSAPTTTSKAPTTK 379
QY 181 EPAPTTTSKAPTTKPEAPTTTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 240
DB 380 EPAPTTTSKAPTTKPEAPTTTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 439
QY 241 PKKAPTTTKEAPATTKEAPTTTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 300
DB 440 PKKAPTTTKEAPATTKEAPTTTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 499
QY 440 PKKAPTTTKEAPATTKEAPTTTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 499
QY 301 PTTTKEAPATTKEAPSTTKEAPATTTSKAPTTTKEAPATTTSKAPTTTKEAPSTT 360
DB 500 PTTTKEAPATTKEAPSTTKEAPATTTSKAPTTTKEAPATTTSKAPTTTKEAPSTT 559
QY 361 PAPPTTKEAPATTTKKAPTTTKEAPATTTKKAPATTTKKAPATTTKKAPATTTKKAP 420
DB 560 PAPPTTKEAPATTTKKAPTTTKEAPATTTKKAPATTTKKAPATTTKKAPATTTKKAP 619
QY 421 TPKKLTPTTPEKLAPTTPEKAPATTPEELAPTTPEEPTPTTPEEAPTTTKEAAPT 480
DB 620 TPKKLTPTTPEKLAPTTPEKAPATTPEELAPTTPEEPTPTTPEEAPTTTKEAAPT 679
QY 481 PAPPTTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 540
DB 680 PAPPTTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTTSKAPTTTKEAPA 739

QY 541 EPTSTSKAPPTPKGTAPPTKEBPATTPKEBPATTPKGTATPKEBPATTPKPKAP 600
D 740 EPTSTCKAPPTPKGTAPPTKEBPATTPKEBPATTPKGTATPKEBPATTPKPKAP 799
QY 601 KELAPTTKGTSTTSKAPPTTPKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 660
D 800 KELAPTTKGTSTTSKAPPTTPKEBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 859
QY 661 PTTTKEPTTHKSPDESTPELSAEPPTKALENSKEPGVPTTKPATKPEMTTAKDKT 720
D 860 PTTTKEPTTHKSPDESTPELSAEPPTKALENSKEPGVPTTKPATKPEMTTAKDKT 919
QY 721 TERLRTPTTETTAAPKMTKETATTEKTESKITATTTGVTSTTODTTPFKITLKT 780
D 920 TERLRTPTTETTAAPKMTKETATTEKTESKITATTTGVTSTTODTTPFKITLKT 979
QY 781 TLAPKVTTKKTTTETINMKPEETAKPKRATNSKATTPPKPKAPKPKPTSTKPKPT 840
D 980 TLAPKVTTKKTTTETINMKPEETAKPKRATNSKATTPPKPKAPKPKPTSTKPKPT 1039
QY 841 MPVRKPKPTTPPKMTSTMPKELNPTSRILAEALQTTTRNPTNSKLVNPKSDAG 900
D 1040 MPVRKPKPTTPPKMTSTMPKELNPTSRILAEALQTTTRNPTNSKLVNPKSDAG 1099
QY 901 ABEETPMMLRPHVFMPEVTPMDYLPRVNOGIIINPMLSDETNINCKRPVGLTTLN 960
D 1100 ABEETPMMLRPHVFMPEVTPMDYLPRVNOGIIINPMLSDETNINCKRPVGLTTLN 1159
QY 961 GTLVARGHFMFLSPSPSPARRITEWGISPIDYTRNCNGSKTFFKDSQYMR 1020
D 1160 GTLVARGHFMFLSPSPSPARRITEWGISPIDYTRNCNGSKTFFKDSQYMR 1219
QY 1021 TNDIKAGYPKPIFKGFGGLTGOIYALSTAKYKNMPEVYEFK 1064
D 1220 TNDIKAGYPKPIFKGFGGLTGOIYALSTAKYKNMPEVYEFK 1263

RESULT 3
09JUN99 PRELIMINARY; PRT: 1054 AA.

AC 09JUN99; 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN PRG4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikegawa S., Nakamura Y.;
RT "a novel mouse gene highly homologous to a human gene encoding
RT megakaryocyte stimulating factor precursor and cartilage superficial
RT zone protein.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB034730; BAA92310.1; -.
DR MGI: 1891344; Prg4.
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR002965; P_rich_extens.
DR Pfam: PF01033; Somatomedin_B.
DR PRINTS: PR01217; PRICHEXTENS.
DR PRINTS: PR00022; SOMATOMEDINB.
DR SMART: SM00120; HX; 2.
DR SMART: SM00201; SO; 2.
DR PROSITE: PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE: PS00524; SOMATOMEDIN_B; 2.
DR SEQUENCE 1054 AA: 115991 MW: 4664BFA2283235 CRC64:

Query Match 44.2%; Score 2519; DB 11; Length 1054;
Best Local Similarity 49.6%; Pred. No. 2,4e-160;
Matches 533; Conservative 38; Mismatches 136; Indels 368; Gaps 20;

QY 1 VKDNKKNRKTKPKRPVVDGSGDNDGDFKVT--PTDITGTHNKVSTSPKTTAKP 58
D 195 VKDNKKNRKTKPKRPVVDGSGDNDGDFKVT--PTDITGTHNKVSTSPKTTAKP 254
QY 59 INRPSPNSDTSKETSILVNEKETTETTTNKOTSTGKEKETSASKESTSEKTS 118
D 255 VTPKPSLANSSESKASLSAKNETTVERKETTAINKQSA-SKKKTSVKEKSAEKTS 313
QY 119 AKDLAPTSVLAKPPTKAPTTPGALTPPKBPPTPKBPASTPKBPPTTIKAPPT 178
D 314 KDVY-----EPTSTPK-----NSAPTT 331
QY 179 PKEAPTTTSAPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 238
D 332 TKKPV-TTKESKFLP-----LPDEEPTTAKBPPTTKBPPTTKBPPTTKBPPTTK 376
QY 239 TTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 298
D 377 TTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPK 436
QY 299 PAPTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 358
D 437 PEPPTTKEESTTRKEBPPTPKBP-----EPTPKBPPTTL 474
QY 359 KEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 418
D 475 KEPPTTTPKBPPTTPKBPPTTPKBPPTTPKBPPTTPKBPPTTPKBPPTTPKBPPTTP 512
QY 419 PTPPKKLPTTPKLPATTPKBPATTPPELAPTTPEBPPTPEBPATTPKAAADNTP 478
D 513 -----TP 514
QY 479 KEPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 538
D 515 KEPPTTTPKBPPTTPKBPPTTP-----KEPPTTPKBP----- 549
QY 539 TKEPTSTTSKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 598
D 550 -----EPTTPKBPVPTPKBPPTTP-----KEPPTTTPKEP 581
QY 599 AKELAPTTTGTSTTSKAPPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 658
D 582 EP-----TTRKEBPPTPKBPPTTPKBPPTTPKBPPTTPKBPPTTPKBPPTTP 617
QY 659 STPTTKEPTTHKSPDESTPELSAEPPTKALENSKEPGVPTTKPATKPEMTTAKD 718
D 618 ----- 617
QY 719 KTTREDLRTPTTETTAAPKMTKETATTEKTESKITATTTGVTSTTODTTPFKITLKT 778
D 618 -----TSPKT-----TTLK 626
QY 779 TTTAPKVTTKKTTTETINMKPEETAKPKRATNSKAT-----TPKP-OKTPKAP 829
D 627 ATTLAPKVTAPAE-----ELONKPEETTPASEDSOSKTLTKQKPTKAPKPKPTKAP 861
QY 830 KKPSTKPKPTMPVRKPKPTTPPKMTSTMPKELNPTSRILAEALQTTTRNPTNSKLVN 889
D 682 KKPSTKPKPT-MPKRPKPTTPAPLKTTSATPELNTTP--LEVMPTTITPKOTPEETA 738
QY 890 EVNPKSEDAGAGETPMMLRPHVFMPEVTPMDYLPRVNOGIIINPMLSDETNINCK 949
D 739 EVNPKSEDAGAGGEGEKP-LTPGPVLEPTAIPGTDLAAGRLNGLININPMPSEDNMLCNG 797
QY 950 KPVVDGLTTLNGLTVAFRGHYFMFLSPSPSPARRITEWGISPIDYTRNCNGSKT 1009
D 798 KPVVDGLTTLNGLTVAFRGHYFMFLNPRPSPARRITEWGISPIDYTRNCNGSKT 857
QY 1010 FFFKDSQYMRFTNDIKAGYPKPIFKGFGGLTGOIYALSTAKYKNMPEVYEFK 1064

QY 676 ESTPELSAPPTKALENSKPEPVPT--KTPAATKPEMTTAKDKTTERDLRTT-----P 729
 Db 718 QTTTTTAPET-----STEPSSTTPVQTTTITAPETSTEPSSTTPVQTTTAP 771
 QY 730 ETTTAPKMTKATTEKTESKATTTATTOVSTTODTTPETTKITKTTLAPKVT 789
 Db 772 ETTSTEPSSTTPVQTTTITAPETSTEPSSTTPVQTT--TITAPETSTEPSSTN 829
 QY 790 KKTITTEIMKPEETA-KPRDRATNSKATTPKPKAKPKPTSKPKPTMTRVKK 848
 Db 830 TPVQTTT--TTAPETSTEPSSTTPVQTTTITAPETSTEPSSTTPVQTTTITAP 887
 QY 849 TTPPRKMTSTMPPELNPSTRALAEALQTTTPPNQTPNSKLEVPKSDAGAGC-ETPH 907
 Db 888 TTSTEPSSTTPVQTTTITAPET--TSTEP--PSSSTTPVQTTTITAPETSTEPS 941
 QY 908 MLRPHVPMPEVTP--DMDLPRV-----PNQIITNMLSDEN----- 945
 Db 942 SSTTPVQTTTTPVQDCLSIDRVVPTTEEMENKRDIIIOGYSPPRTAJSVSTFD 1001
 QY 946 -----ICGKRPVQDGLITLRNGTLVAFNGHYFMLS-----PSPSPARITTEW 990
 Db 1002 IGTCTAILCTYTSEGGISNL-NATL-----FIGLSDSSIDLPE-----Y 1041
 QY 991 GIPSPIDVFTRCNCEGKTFEKKSOQVWRETN 1022
 Db 1042 NVPTGLE-IMPEINCEGNK-----WSTNN 1064
 RESULT 6
 Q917S1 PRELIMINARY; PRT: 1049 AA.
 ID Q917S1 PRELIMINARY; PRT: 1049 AA.
 AC Q917S1;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE CG5228 PROTEIN.
 GN CG5228.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Baau A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
 RA Borovica D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng S., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunnov B.C., Dunn P.,
 RA Durlin K.J., Evangelista C.C., Ferraz C., Ferreira S., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibovnam C.,
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodali S., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slater E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sylvestre R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhan G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003495; AAC22353.1; -;
 DR FlyBase: FBgn030561; CG5228.
 SQ SEQUENCE 1049 AA; 107278 MW; 954DD629E7619671 CRC64;

Query Match 16.4%; Score 936.5; DB 5; Length 1049;
 Best local Similarity 34.6%; Pred. No. 1e-54;
 Matches 371; Conservative 45; Mismatches 439; Indels 217; Gaps 54;

QY 2 KDNKKRRKKKPKRPVYDEAGSGLDNGDFKVTTPDTS---TQHNKVSPTSPTTAKP 58
 Db 30 KNEIGTTAKPTTLKP-----TEGTSAKPTTLKPTEGTSAKPTTLKP 71
 QY 59 -----INPRSLPMSDSKETSLS-----TVNKEVTVETKETTNNKQF--STDGKERT 106
 Db 72 TEGTTAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEG 127
 QY 107 SAKET--OSIEKTSAK--DLAPTSKVLAKPT-----PKAETTKGALTTKEPPTT 155
 Db 128 SAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTSAKPTTLKPTEGTSAKPT 187
 QY 156 PKE-----PASTPKEPT---PTTISAPTPKKEPAPPTTKSAPTPKKEPAPPTTK 207
 Db 188 PTEGTSAPPTTLKPTEGTSAPPTTLK--PTEGTTAKPTTLK--PTEGTSAPPTTLKPTEG 243
 QY 208 TTPKKEPAPPTTKKEPAPPTTKSAPTPKE--PAPTPKKEPAPPTTKKEPAPPTTK 262
 Db 244 TTAK---PTTLKPTEGTTAKPTTLKPTEGTSAPPTTLKPTEGTTAKPTTLKPTEGTS 300
 QY 263 TTPKKEPAPPTTKKEPAPPTTKKEPAPPTTKKEPAPPTTKKEPAPPTTKKEPAP 317
 Db 301 TTKLPTEGTTAKPTTLKPTEGTSAPPTTLKPTEGTTAKPTTLKPTEGTSAPPTTLKP 360
 QY 318 TTPKKEPAPPT---TKSAPPTTKKEPAPPTTKSA---PTTKKEPSPPTTKKEPAPPTKE 368
 Db 361 TTAK---PTTLKPTEGTSAPPTTLKPTEGTTAKPTTLKPTEGTSAPPTTLKPTEGTTAK 416
 QY 369 PAPTPPKKAPPTTKKEPAPPTTKKEPAPPTTKKPA---PTAPKKEPAPPTTKETAPTT--PK 423
 Db 417 --PTTLKPTEGTTAK---PTTLK--PTEGTTAKPTTLKPTEGTSAPPTTLKPTEGTTAKPT 470
 QY 424 KLTPTE-----TPELATPTEKAPPTPELATPT--TPEPTPTTPEEAPPTPKAAPN 476
 Db 471 TLKTEGTTAKPTTLKPTEGTSAPPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTE 530
 QY 477 TPKEPAPPTTKKEPAPPTTKKEPAPPTTKETAPTTKGTA-----PTTLKEPAPPTP 526
 Db 531 KGTAKPTTLK---PTEGTSAPPTTLKPTEGTTAKPTTLKPTEGTSAPPTTLKPTEGTTA 587
 QY 527 K-----KKA-----PKLATP-----TKKEPTSTTSOKPA-----PTTKPGKT 558
 Db 588 KPTTLKPTEGTSAPPTTLKPTEGTSAPPTTLKPTEGTTAKPTTLKPTEGTSAPPTTLKP 647
 QY 559 APPTPKKEPAPPTTKKEPAPPTTKGTAPTTLKEPAPPTPK---KPAKELAPPTTKPGPT 614
 Db 648 EGTAK---PTTLK---PTEGTSAPPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEG 701
 QY 615 TSKPAPPTPKKE--TAPTPKKEPAPPTTKKPA---PTPEPTPPTTSEVSTPT--TKKEPT 668
 Db 702 TSKAPPTTLKPTEGTTAKPTTLKPTEGTSAPPTTLKPTEGTTAKPTTLKPTEGTTAKPT 761

QY 669 TIHSPDSTBELSAPPP-----KALENSPEPGV--PTTKT---PATK 709
 DB 762 TL--KPTGCT--SAPTKLPTBGTAKPTTLKPTBETSAPTKLPTBETSAPTKL 816
 QY 710 PMTTAKDKTERLRLT-----PETTAAPKMKETATT-----TEKTES 752
 DB 817 PTEGTAKPTTLKPTBETSAPTKLPTBGTAKPTTLKPTBETSAPTKLPTBGTAK 876
 QY 753 KITATTOVSTTTODTTPFKITLTKTTAPKVTTK--TITTELMNKDEETAKPD 810
 DB 877 PTLKPTBETSAPTKLPTBGTAKPTTLKPTDGTAKPTTLKPTBETSAPTKLPTB 936
 QY 811 RANPNTATPKPOKPTKAPK--KPT--STKPKTMPRVKPKTPT--PRKMTSTMP- 861
 DB 937 -GTAKPTTLKPTBETSAPTKLPTBGTAKPTTLKPTBETSAPTKLPTBETSAP 995
 QY 862 ELNPTSRIAEAMLOTTTPRNPQNSKLVENPKSEDAGAGS-ETPHMLIR 912
 DB 996 TLKPTERTSAO--PTTLKPTBGTAKPTTLKPTKGTSGANNEETKRERRR 1045

RESULT 7

ID 076894 PRELIMINARY: PRT: 1795 AA.
 AC 076894:
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE EG:5667.1 PROTEIN.
 GN EG:5667.1 OR CG14796.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; Pubmed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anagnostis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaziel R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-Y., Andrews-Plamkoc C., Baldwin D.,
 RA Bellow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dudin K.J., Evangelista C.C., Feriz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellan A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Jeggam C.,
 RA Jaitali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson J.,
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,
 RA Palazzolo M., Peltum G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svatskis R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
 RT "The genome sequence of Drosophila melanogaster".
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cadieu E., Dreano S., Lejaune V., Mottier S., Gallibert F.,
 RT Sequencing the distal X chromosome of Drosophila melanogaster";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Benos P.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AEO03421; AAF5644.1; -;
 DR EMBL: AL031028; CAI19845.2; -;
 DR FLYbase: FBgn0025390; EG:5667.1
 DR InterPro: IPR002557; Chitin_binding.
 DR InterPro: IPR002965; P_rich_extensn.
 DR Pfam: PF01607; Chitin_bind_2; 2.
 DR PRINTS: PR01217; PRICHEXTNSN.
 DR SMART: SM00494; ChIBD2; 2.
 SQ SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match 15.2%; Score 865.5; DB 5; Length 1795;
 Best Local Similarity 28.5%; Pred. No. 9,2e-50;
 Matches 276; Conservative 129; Mismatches 447; Indels 115; Gaps 28;

QY 6 KNRRKKKPPPPVVDGSGLDNGDFVYTPDPDSTQHNKVNSTPKITAKPINPSSL 65
 DB 431 KNDVTEAPEIKSPL-----KGILHSENIVILPEYTTT--TTTTKPVVLICPTISPDIT 483
 QY 66 PINSPT-----SKETSLVYNKEVETVEETTTNKGSTDEKETSKEQSI 114
 DB 484 PPSSTTAVATSKPISSTEDHSTTAKTTTKRTVTEKTSATEKPRVTVTTTQK 543
 QY 115 EKTSAKDIAPTSKVLAKPPKPAETTTKGPALTTKEPPTTPKPEASTPKPEPTIKS 174
 DB 544 RSTTHNHSPTDKTIRSTLSKPTTTPSTTPSTTPST-----TTPSTTPST--- 594
 QY 175 APPTKEPAPPTTKSAPPTPKPEAPPTTK-EPAPPTKEPAPPTTKPEAPPTTKSAPT 233
 DB 595 --TTPSTTPSTTTTKVSTHPRRTSOKTTTASTTK-----TTPSKTKTDIDPTST 648
 QY 234 KEAPPTPKPAPPTPKPEAPPTPKPEPPTTPKPEAPPTTKPEAPPTTKPEAPTKKAP 293
 DB 649 TSKLSTTQKTTTTHKFTAAATSTSEKPTTEKSTVSTTKKSTESSPRTSSTGKPT 708
 QY 294 TTPKEPAPPTPKPEAPPTTKPEAPPTTKSAPPTTKSAPPTTKPEAPPTTKSAPTREP 353
 DB 709 TTPKEPAPPTTKKVTITTOITTTTPLKSSSTETTSQPTTTTPQPTTTTLVTPKTS 768
 QY 354 SPPTKEPAPPTPKPEAPPTPK-----KPAPTKEPAP-----TTPKE 392
 DB 769 TTTTTEKPTTSPKPTTOKTTSTAPNTKVAITTOKEETPTPOSTSTTPTTKTTNN 828
 QY 393 PAPTTPKAPPTKAPPTTK-ETAPPTPKKLTPTTEKLAAPTPEK-----AP 443
 DB 829 PEPTSTETPTSTTPKPSSTTPKTSVASTETKTTSSKPTTEKSTENPNSVTSAL 888
 QY 444 TTEBELAPTPEEPPTTTEEPAPTTPKAAAPNTKEPAPTTPKEPAP--TPKEPAPT 501
 DB 889 TTSQKATSTSPKTT-QNITTTTPKPTTEKTSQEQATTSOKVSTVITTTKATESS 947
 QY 502 PKETAPPTPKGAPPTLKEPAPPTPKKPAKELAPTTTEPT--STSDPKAPPTPKGA 559
 DB 948 PLTLSTEEBNTPKPLKTTTTSVATTRITTTTISSESTETTSOKPKSTPTTST 1007
 QY 560 PTPKPEAPPTPKPEAPPTPKGAPPTLKEPAPPTPKKPAKELAP--TTKGPTSTTSK 618
 DB 1008 RTTPKVTIVISQNTPTTTSKSTVIT-----TTP-NPSPSQBRPTTTRQPTSTAST 1061
 QY 619 PAPTPEAPPTPKPEAPPTPKKPAPTTPEPTPTTSEKSTTTT-TKE----- 666


```

Db 1062 TSGTTRIPPTTNPNSSSTDLTNTVRPPCPDDSDSKNTNACQGLQVNLLEQS 1121
QY 667 -----PTTIHKSPODESTEPLESAEPKPALENSPKPEPVPTTKTAATKPEKTTAKD 718
Db 1122 PQKQDQTHHTHTALGSRNTLGGQEVDPYMDAPSSAENESQAAVTAAAPMTSLAAA 1181
QY 719 KTRR---DLKTPBETTTAAAPKMTKETATTEKTESKITATTTQVSTTQDTPPKIT 775
Db 1182 HLQKLFHIISTTPSRREHAP--TQRPSSQPSQSRN-CVITIAQMRNHLATSKPIAH 1238
QY 776 TLKTT-----TLAPKVTTKTKTITITTEIMKNPEETAKPKDRATNSKATPPQKPTKA-PK 830
Db 1239 SLKLSIQQLASTQKRSIPPKTLVTHNTKPED-SEYDSESEQYTDENEVLDTQPR 1297
QY 831 KPTSTKPKTMPVRKPKETTPPKMTSTMPBELNPTSRIAEMLOTTTRNPOTNSKIYE 890
Db 1298 AMSSTVAALVPAVPSTTTEREPQK-TSSSP--SPT-----KATSTTTOPIETTTGDLXY 1350
QY 891 VNPKSED 897
Db 1351 DSGSSD 1357

```

```

RESULT 8
O96449 PRELIMINARY: PRT: 1489 AA.
AC O96449:
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DE 01-MAY-1999 (TREMblrel. 10, Last annotation update)
DE CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
GN CARO.
OS Euphorbia infestans (Potato late blight fungus).
OC Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC Phytophthora.
OC NCBL_TaxID=4787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RACE 1-11;
RA Goernhardt B.;
RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF061185; AAC72308.1;
SQ SEQUENCE 1489 AA; 164037 MW; 764CC97D1C2F5163 CRC64;

```

```

Query Match 15.0%; Score 853; DB 10; Length 1489;
Best Local Similarity 32.38; Pred. No. 5.3e-49;
Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;

```

```

QY 9 TKKKPPKPPVVDAGSLDNGDFVYTPPDSTT-----QHNKVSIPKITTAKPKNR 62
Db 338 TRYAPTEKPYDVEETTYTTEESTYAPTKSEINAPTERMHVHIEPCDTEVTMYAPEET 397
QY 63 PSLP-----PNSDT-----SKETSLTNKETTVEKETT---TNKQSTSD 100
Db 398 TYAPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 457
QY 101 GKEKTSKAKETOSIEKTSKADLAPSKVLAKPKAKETTTGSLITTPKEPTPTTPEPA 160
Db 458 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 514
QY 161 STTPEPTPTTIKSAPTTPKPAPT-TTKSAPTPKPEAPT--TTKBPAPTTPKPAPT 217
Db 515 EETPEPEPEET-TYAPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 573
QY 218 TKBPAPTTTKAPT-----TKBPAPTTTPKKAAPTTPKBPAPTTPKEPTPTTPE 267
Db 574 TYAPTEETTYAPEETMYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 632
QY 268 P-----AP-----TKBPAPTTPKEP-----APAPKKAAPTTPKBPAPTTP 304
Db 633 PTEETTYASTTEETTYAPTEETTYAPAEETTYEPPEETTYAPTEETTYAPTEETTYAPTE 692

```

```

QY 305 KEAPAT--TKKPSPTTPKEP-----APT--TKKSAPT-----TKKEP-----A 339
Db 693 TYAPTEETTYAPAEETTYEPPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 752
QY 340 PT-TTKSAPT--TKKPSPTTPKEP-----APTTPKBPAPTTPKKAAPTTPKBPAPT-- 388
Db 753 PTEATTYAPTEETTYAPAEETTYEPPEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 812
QY 389 TKBPAPTTTKKPAATAPKPAAPTTPKETAAPTTPKLLPTTPEKAPAPTTPKPAPTTPEE 448
Db 813 TPEPEPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTY 870
QY 449 LAPT--TPEEPPTTPEEP-----APT-----TPKAAAPNTKBPAPTTPKEP- 489
Db 871 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 930
QY 490 -----APTTPKBPAPTTPKETAAPTTPKGTAPT--TKBPAPTTPKKAAPK-ELAPT--TT 539
Db 931 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETTYEPPEET 990
QY 540 KEPTSTSDKPAATTPKGTAPTTPKBPAPTTPKEP-----APTTPKGTAPT----- 585
Db 991 YAPTEETMYAPTEETTYAPTEETTYAPAEETTYEPPEETTYAPTEETTYAPTEETTYAST 1050
QY 586 -----TLKEAPATTPKAP-----APKE---LAPT--TKGPTSTSDKPAAPTTP 624
Db 1051 EETTYAPTEETTYAPAEETTYEPPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1110
QY 625 KETAPTPKBPAPTTPKPAAPTTPETPPPTSEVSTPTTPKEP--TIHKSPODETELS 682
Db 1111 YAPAEETTYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1169
QY 683 AEPPKPA-LENSPKEPC-----VPTTKPATKPEMTTAKDKTTERDLTTP---ETT 732
Db 1170 TEETTYAPTEETTYEPGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1229
QY 733 TAAPKMTKETATTEKTESKITATTTQVSTTQDPT-----PKTIIT--LKT 779
Db 1230 TYAP-----TEETTYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 1284
QY 780 TTLAPKVTTKTKTITTEIMKNPEETAKPKDRAT--NSKATTPKP-OKPKAKKPP--- 833
Db 1285 TYAPTEATTYAPTEETTYAPTEETTYEPGETTYAPTEETTYAPTEETTYAPTEETTYA 1344
QY 834 -----STKKP-KMPVRKPKETTPPKMTSTMPBELNPTSRIAEMLOTTTRP--N 881
Db 1345 PAESTYSTETKPCNTEETTEDPTDE--PSDEPTDEPTDEPTDLPTEDEPTPCDN 1402
QY 882 QTPNSKLVEVNPKSEDAQ 899
Db 1403 GINGIGIVENKVRNNAG 1420

```

```

RESULT 9
O9SPMO PRELIMINARY: PRT: 1315 AA.
AC O9SPMO:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE EXTENSIN-LIKE PROTEIN.
GN PEX2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta, Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
RN NCBL_TaxID=4577;
RP SEQUENCE FROM N.A.
RC TISSUE=POLLEN;
RA Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
Hohorst D., Gao M., Showalter A., Bedinger P.A.;

```



```

Db 274 KNPETKKKPEWDETPVEVEKPEPVKEKAPVLKKKKDPAKADPSPSKAAPKPEVS 333
Qy 51 PATTAKPI-NRPSLPNDSSTKETSLSVKNKETVETKETTITTKQSTGKEKETSAAK 109
Db 334 SPVVPPTPVKNPVKKYKRPWEDEDPABEVEKPSAPKEKTVLKKKEPSSSTPSSPS 393
Qy 110 EQOSTEKSADLAFTSVLAKPKPKAE---TTTKGA-----LTPP 148
Db 394 PKKAPPAVKPRDSSPKKATPIQADBPKAQEVPPPTPVKNPVKKYKRPWEVDEDDPVEVKOP 453
Qy 149 KEPPTPTP-----KEPASITPEKPE---TP-TTISAPITPEKPAITTKASAPITPEKPA 199
Db 454 EAPAKKKTIVLAKKEPAADOTAPKATSKTPEPEKKDPVKPRDSSPKKVAAPDSQAQAPA- 512
Qy 200 TTTKEPA-----PTTKEPAITTKEPAP---TTKSAPITP 233
Db 513 TPVKNPVKKRPWEDETPADVSKPTDAKKTPSLAKKODAPAKESLAKRADTKAAKAP 572
Qy 234 KEP-----APITPKKPAITTKPE---PAPITPEKPTTTPKEPAPITTKEPAPITPK-EP 283
Db 573 RDPSPKVAAPAPKKTTPVLAKKEPAPADSKTKEPEKSKRPDPSPKKAVPAKPVKTEV 632
Qy 284 APAPKPKPA-----TTKEPAPITTKEPAPITTKEP---SPTTKEPAPITTKSAPIT 333
Db 633 APAAVKKPEPISKPDOTAPKKAEPNSVVP-PTPVKNPVKKRPWEDDAPAKPVSLPE 691
Qy 334 TTKPAPITTKSAPITP-KEPSPITTKEPAPITPK-EPAPITPKKPA---TTPEKAPIT 388
Db 692 PEK-TVLVAKKAPTKDSEAAADPVSGPSSDKPLAKKAPRDPSPKAPVIRKAPK 750
Qy 389 TPKEPAPITTKKPAAPTA-PKEPAPITTKETAITPKLITPTP-----EKLAPITPEKPA 442
Db 751 T--EVPAAVKKKEPVAKSDPSKAK-AEPNSP--VPPPTPVKNPVKKRPWEDDDA 805
Qy 443 PTTPEELAPITPEEP-----TPTTPEPAPITTKAAAPNTPKEPAPITTKEPAPITP-PKE 496
Db 806 PABEVNAPPEBEKTPVLAKKTPVKRDPSPSKAVPAKESKITDAPVSVAKKPEPVSKPE 865
Qy 497 PAPITPEKAPITPKGAPITTKLEPAPITTKAPKELAPIT-TTKEPSTTSODKAPITP 555
Db 866 PSKKAKEPNSVVP---PTPVKNPVKKW-KPWEDEDETEVEKKPSE--DEKKTPLVA 918
Qy 556 KGTAPTPKPAPITPKPAPITPKGTAPTTKLEPAPIT-----TKKPA---KEL 603
Db 919 K-KEPEKPD-APKVAKRPDPSKAVE--KEPAKVAKPRDLSFKAIPIPANTOA 974
Qy 604 APITTKGPS-----TSDKP-----APITPKET-----APITPKPAPITPKKAPIT 646
Db 975 PPTPVKNPVKKRPWEDEDEPAEVSAPPEKTPVLAKKAPAKRDP---SPKKAAPV 1031
Qy 647 TPETPPTSEVSTPTTKEPTTIHKS-----DETPELSA-EP---TPKALENSPKERG 698
Db 1032 AAK-PDPKILEP-PTPVKNPVKKRPWEDEDESEPAEPPKAPITVLAKKAPITKPA 1089
Qy 699 V-----PTTKTAPAT-----KPEMTTAKDTERDLRTTETTT--TAAPK 737
Db 1090 TRPDSEAAADPVSGTSPKLSKKAAPVEKPKPTDPODKLAKPAPAKKEAPAPAPAK 1149
Qy 738 MKKEPATTEKTESKITATTTQVSTITQDITTPKITTLLKTTLAPVTTTKTITTE 797
Db 1150 KKKFVWDDPDEPDAFTVPAPSKRPDEDPAPLG-----GPKTKDPK----- 1193
Qy 798 INNKDEETAKPKDRAATNSKATTPKPOKPTKAPKKTSTKKPKTMRVAKPK----- 848
Db 1194 -LNKKAAPAKPTKE-----PKKEVSKPEPKTEPPKP-AAPKWKRPWEDDPDEPE 1243
Qy 849 ---TTPTPKKMTSTMP 861
Db 1244 ADFTWAPAKKPDTEDP 1259

```

RESULT 11
Q9LIB8

```

ID Q9LIB8 PRELIMINARY; PRT: 1480 AA.
AC Q9LIB8;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINER PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety pl.
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL: AP001306; BAB03062.1; -.
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PRO1217; PRICEXTENS.
SQ SEQUENCE 1480 AA; 147153 MW; D1AC0C79F155E732 CRC64;

```

```

Query Match 14.0%; Score 800.5; DB 10; Length 1480;
Best Local Similarity 29.6%; Pred. No. 1,76-45;
Matches 307; Conservative 54; Mismatches 505; Indels 171; Gaps 45;

Qy 10 KKKPKRPVNDGAGSLDNGDFKVTPTDSTQHNKYSTS-----PKTTAKPI-NRPSL 65
Db 117 KRPAKRPVNVK-----PHNPKRPTKRNHPRKPTKNNHPRKPTKPKKPPSV 167
Qy 66 PNSTSKESTLSVKNKETVETKETTITTKNOTSNDGKEKTSAAKETSIEKTSADLAFT 125
Db 168 KPPSTPK-----PPTNPPSPHPQRPETHKRPSCPTTPPAVSPNMTTP 211
Qy 126 SKV--LAKPTKAEITTKGALTPPKERT--TTP--KEPASTPKERTPTTKSAPITP 179
Db 212 TQMPRIAP-PIASPVATPIATPPIATPPIIPVATPPIITPPIANPPIIMPPIATP 270
Qy 180 KEAPITP-----KSAPITPKPAPITP-----KEPAPITPKPAPITTK 220
Db 271 PVAAPPIINPISKRPVITP---PTTPIAKPIATPPISTPPATTPPAATPPIITLP 326
Qy 221 PA-----PTTTSAPITP-KEPAPITPKKPAV--TTPEKAPITPKERTPTTKEPAP 270
Db 327 PAKPPVAISPLVT--PVTPIAOPVATPPIATPPIATPPIATPPIATPPIATPPISESP 384
Qy 271 TTKPAPITP-KEPAPAPKPAAPITPKPAPITPKPAPITP-----KESPTTKEPAP 325
Db 385 VATPPTATSPKIPKPPAPKAPVATPPIAKSPIATPPIATPPIATPPIAIPKPPVATPPTTPP 444
Qy 326 TTTTSAPITTKEPAPITTKSAPITPKEP--SPTTKEPAPITTKEPAP--TTPKKPAPT 380
Db 445 TAT--PVAAPVEMPIPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPI 501
Qy 381 TPKEPAPITPKEP--APITTKKPAAPTA--KEPAPITPKETAITP-----PKKLTTP--T 428
Db 502 TPPLATPVAAPKPPVVPPIATPPIATPPIAKSSVATPPIATPPIATPPIATPPIATPPI 561
Qy 429 TPKEAPITPKKAPITPEELAPIT--PEEPTPTPEPAPITTKAAAPNT--PKPAPIT 484
Db 562 TPPLATPVAAPKPPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPI 621
Qy 485 TPKEPAPITPKKAPITPKETAP--TTPGTAITP--LKEPAPITPKKAPKELAPITP 539

```

```
Db 622 TPPIATPPIAKPPVATPPATPPATSPVAKPPVATPPIKPPAKPPVATPPATPPVA 681
QY 540 KEPTSTSDKAPATTPKGTAP--TTKEEAPATTPKEEAPATTPKGIAPTLIKEAPATTPKKP 598
Db 682 KRPVAT---PPTATPPATPPATPPVATPPATPPVATPPATPPATPPATPPATPPVA 736
QY 599 AKELAPATTPKGTSP--TSDKAPATTPKEATPT--TPKEEAPATTPKKAPATTP-----E 649
Db 737 MPPIATPPAKPPATPPATPPATPPVAKPPVATPPATPPATPPATPPATPPATPPVA 796
QY 650 TPEPTSEVSTPTTKEPTTIIH--KSPDESTPELSAEPKALENSKEGVPTTKTPAA 707
Db 797 TPPIAKPPVATPPATPPATPPVAKPPVATPPATPPATPPATPPATPPATPPATPPVA 856
QY 708 TKDEMT--TTAKDKTTERDLKTPEETTAAPKMKETATTEKTESKIFATITQVSTSTT 766
Db 857 TPPIATPPAKPPVATPPATPPATPP--IAKPPVATPPATPPATPPATPPVAKPPVATPPATPP 914
QY 767 QDTTPKITTLKTTTLAPKVTYTKKT---ITTEIMNKPETETAKKDRATNSKATTPKP 822
Db 915 PAKPPVATPPATPPVAKPPVATPPATPPATPPATPPATPPVATPPATPPATPPVA 971
QY 823 QKPTKAPKPTSTKKPKMPVRKPKTTTPPKMTSTMBELN-----TSRIAE--- 871
Db 972 KPPPTTP--PTAT--PPVAMPPIATP--PPAKPPVATPPATPPVAKPPVATPPATPP 1025
QY 872 ----ANLQTTTPNOTPNKSL--VEVNPKESEDAGACETPPMLLRPHVMEVTPDMD 924
Db 1026 VLPIIAKPPVETSPATPPATPPATPPVAIRP-----VVKPPVATPPATPPVA 1070
QY 925 YLPRVNPQGIINPMLS 941
Db 1071 ATPVNTNPATAMPPIVT 1087

RESULT 12
Q41805 PRELIMINARY; PRT; 1188 AA.
ID 041805
AC 041805;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE EXTENSIN-LIKE PROTEIN PRECURSOR.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B73; TISSUE=POLLEN;
RA Rubinstejn A.L.; Broadwater A.H.; Lowrey K.; Bedinger P.A.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z34465; CAAB4230.1; -
DR Mendei; 14346; Zeama; 2368; 14346.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR Pfam; PF00560; LRR; 3.
DR SMART; SM00370; LRR; 4.
KW signal.
FT SIGNAL.
SQ SEQUENCE 1188 AA; 120980 MW; 2C77C7F8D7130149 CRC64;
```

Query Match 13.8%; Score 784; DB 10: Length 1188;
Best Local Similarity 29.9%; Pred. No. 1.7e-44;
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

```
QY 13 PPKPPVAVVEAGSGLDGDPKVTTPDTSTQNNKSTSPKITTAKPINRPSL-----PPN 68
Db 460 PTPHSPAD-----DYVPTTPPVKSPKSPATSPSPQVPPAASRTPPELVKLSPPQ 510
QY 69 SPTSKESTLVNKEETTVEETKTTTNKQTSDDGKEKTSAKETQSIEKTSAMDLAPTSIV 128
```

```
Db 511 APVG-----SPPPVKTTSPAPIG-----SPSPPPVSVV 541
QY 129 -----LAKPTKATTTKGPALTTPKEPTPTTPKEAPSTTPKEEPTTTKSAPTTKEPA 183
Db 542 SPPPVKSPPPAVVSGSPPPEKSPPPAPVASSPPVSKSP--PPPLVASSPPPVKSSP 599
QY 184 PTTKSAFTTP--KEPATTTKEPATTTTPKEPATTTTKEAPATTTKSAFTTP--KEPAT 240
Db 600 PPAPVASSPPPVKSPPPPTPVASPPPPAPVASSPPPKSPPPPTPVASSPPPEKSPPPP 659
QY 241 PKKAPATTPKKEAPATTT-----PKE-----PPTTPKEPATTTKEPATTTKEPATAPK 291
Db 660 PAKSTPPEEPTPTPTTSKSSPPPEKSLPPTLLSPDQKPTPSTPSKP--PSSPEK 718
QY 292 APPTKEEAPATTPKEPATTTKEPSPTTPKEPATTTTSAFTTTKSAFTTTKEPATTTKSAFT 351
Db 719 SP--PKEVSSPPQTP-----KSSPPPAVSSPPTPVSSPPALAPVSSPSVSKSP----- 768
QY 352 EPSPTTKEPATTTKEPATTTPKKAPATTTKEPATTTKEPATTTTTPKAPATTPKKEPA 411
Db 769 PPAPLSSPPAPQVKSPPPVQVSSP--PPAKSSPPLAPVSSPPQVKEKTSPPAPLSSP 827
QY 412 TPKKETAP-----TRP-----KLTPTTPKLAFTTPKEKPA-----TPEELAPTTPEEPT 458
Db 828 LAKSSPPHVAVSSPPVAVKSSPPAPVSSPPLTPKASPAHVASSPPEVVKPSTP--PA 885
QY 459 PTTPEEAPATTTKAAAPNTPKK--PATTPKKEPATTTKEPATTTKEPATTTTPKGIAPTT 517
Db 886 PTTVISP--PSEPKSSPPPVVSLPPIVYKSSPPPMVSSP--PMTPKSSPPPVVSSPPT 943
QY 518 LKE---PA-----PTTPKK--PAKELAPTTKE--PTSTSDKAPATTPKGIAPTT 563
Db 944 VKSSPPAPVSSPATPKSSPPAPVNLPPPEVKSPPPTPVSSPPA---PKSSPPAP 1000
QY 564 -KEPATTTKEPATTTPKGIATTTLKEPATTTPKKAPKELAPTTTGTSTSDKAPAT 622
Db 1001 MSSPPPEVKSPPPPAPVSSPPPVKSSPPPAVSSP---PPVKSPPPPAVSSPPPV 1057
QY 623 -PKKETATTPPKKAPATTTKKAAPATTTETPTTSEVSTPTTKKPTTIKSPDESTPEL 681
Db 1058 KSPPPAPLSSPPPVKSSPPAPVS--SPPEPVKSPPPAPVSSPPPKSPPPAPVS 1115
QY 682 SAEPTPKALENSKEGVPTTKTPAATKPEMTTAKDKTTEEDLTPTTPT 731
Db 1116 SPPAP-----VKKPSLP--PPAPVSSPPPVTPAPPKKEOSLPPAPS 1158
```

```
RESULT 13
P70670 PRELIMINARY; PRT; 2187 AA.
ID P70670
AC P70670;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
DE MUSCLE-SPECIFIC FORM GP220).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96312450; PubMed=8698236;
RA Yarov W.V.; St-Arnaud R.;
RT "Differential splicing in of a proline-rich exon converts alphaNAC
RT into a muscle-specific transcription factor.";
RL Genes Dev 10:1763-1772(1996).
DR EMBL; U48364; AAB18734.1; -
DR EMBL; U48363; AAB18732.1; -
DR MGD; MGI:106095; Naca.
DR InterPro; IPR002715; NAC.
```

DR InterPro: IPR003037; TS-N.
DR Pfam: PF01849; NAC; 1.
DR Pfam: PF02094; TS-N; 1.
SQ SEQUENCE 2187 AA; 220599 MW; 003646A864DEBFD CRC64;

Query Match 13.6%; Score 776; DB 11; Length 2187;
Best Local Similarity 27.7%; Pred. No. 1,le-43;
Matches 288; Conservative 122; Mismatches 389; Indels 240; Gaps 49;

```
OY 34 VTTPPTSTQHNKSTSKITAKINRPSLSPNSDTSKE-----TSLTVNKETT 84
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 806 VQSPKVDIMSDVPTSPKTSATAV-PK-----DTSATLSKSPAVATLSLSPKAPV 857
OY 85 VETKETTNNKQSTDGKEKTSKETOIEKTSKADLAPTSKVLAKTPAEETTGCA 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 858 APSNATIVPTPEITSLKNALAAAPKKTLLTSIPKTSFS-----PQKTPKSVSLKAPA 913
OY 145 LTT-----PK-----PTPTPKPEASTT-----PKEPPTT 171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 914 MTSKKAIEIAKDVSPSQPFKEVPLLOHVPTSPKSPVSDTLSCALTSPPKGP-PAT 972
OY 172 IKSAPTPPKP-PAPTTTKSAPTTK-----EPAPTTKEPAPTTKEPAPTTKEP 221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 973 LAEPPTVKKSPKPAASAKTIPATPSPEGVAVPLEIPPCSKAKAPTAAPKESSTSSKR 1032
OY 222 APPTT-TKSAPT-----TPKEPAPTTPKK-PAPTTKEPAPTT 257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1033 APKTAVSKEIPSKGVAVPLEISLPKLETSKATPGEKSSAPKRSKTGKPE---TPP 1089
OY 258 KEPTPTPKPEAPTTKEPAPTTKEPAPTTAKKAPTT---PKEPAP---TTKEPAPTT 311
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1090 GGVTAVPPELSLPKPEPQNAATPNESIAASSOKRSPKTSVPKPEPQGVAMLEIPISAP 1149
OY 312 TKESPTPK-----EPAPTTKSAPTTKEPAPTT-KCAPTTPK-----EESPT 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1150 OKAPKTAVPKQIPTEDAVVILAGSPLSPKAKKAPTAAPKAPTPSGVAVANGELISPT 1209
OY 358 TKEPAPTTKEPAPTTKAPTT---PK-----PAPTTKEPAP---T 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1210 KTSKTAAPKENSATLPKSPKTAAPKAPTPATPSSEGVAVPEISPSPTPASKGVPT 1269
OY 397 TTKKPAPTAKEPAPTTKETAAPTTPKLPTPEKLAPTTPPK-PAPTTPELAPTTPE 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1270 LTPKGAANALAE-SPASPKVPTAPAEESTTP-----SPQKIKVAGPKASATPPS 1322
OY 456 EPPPTT-----PEEPAPTTKAAAPNTKEPAPTTKE-PAPTTKEPAPTTKETA 506
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1323 KTKPKTAVPKETSAPSEGVAVPLEIPSPRKAPKTAAPKEMPAPS--PEGATTAAPVQIP 1380
OY 507 PTPPKGAPPTTLKEPAPTT-----KKPAKELAPTTTKEPTS-- 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1381 PSRKSKAKAGSKE-TPTPSPSEGVTAAPLEIPISSKKTSKMAKSPKELTVTPSSKLSQT 1439
OY 545 ----TSDRPAPPTPKGTAPTTKEPAPTTKE-PAPTTPKGAPPTTLKEPAPTTPKKRA 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1440 VGPKESTLGCATVAPLEIPSPKHKAKTYDQVPLTPPK-DAPTTLALE-SPSSSKA-A 1496
OY 600 PKELAPTTKGTSTSDKAPATTKETAAPTTKEPAPTTPKKAPATTPEP-----PPT 654
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1497 PKTAAPPSER-VTVVPEKPA-TPOKASCTASKVPVAETOEVAVSSHRETVTDAVPV 1554
OY 655 TSEVSTPTTK-----EPPTIHKSPDESTPELSAEPKALENSPKPE-PGVPPTTKTAA 707
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1555 KNPSSHKTSKTIELKENATILPSPKTSKPKLSSKKAAPT--SAPKEPASPSTIK-EVT 1611
OY 708 TKPEMTTKKDKTTERDLTTPPETTAAPKMTKETATTEKTESKITATTTQVSTTQ 767
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1612 T--SLAQTAPSLQKAPSTTIKENLAAPAV---LPVSSKSPAPAPASASLSAPATAAP 1665
OY 768 DTTPKPTITLTKTTTAPKVTTKKTIITTEINNKPEPAKRDRAATNSKATTPKQKPTK 827
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1666 QNAPKREATTIIPCKKAAATETIETSTAPSLGAPKETSE-----TSYSKVLMSPP----- 1716
```

```
OY 828 AAKPPTSKTKKPTMPRVKPKKPTTPPKKMTSTMPELNPTSRIAEAMLOTTTBNOTPNK 887
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1717 -PKKASSKRSATLP-----ATILPSLKEASVLS-----PTATSSGK 1752
OY 888 LVEYNPKSEDAAGAGEGTP 906
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1753 DSHISPVV-DACSTGTTTP 1770

RESULT 14
O9VEL9 PRELIMINARY; PRT: 2112 AA.
AC O9VEL9;
DT 01-MAY-2000 (TREMBLrel, 13, Created)
DT 01-MAY-2000 (TREMBLrel, 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel, 17, Last annotation update)
DE CG4090 PROTEIN.
GN CG4090.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins K.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Caddle E., Center A., Chandra I.,
RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Hochs S., Dunkov B.C., Dunn P.,
RA Dubin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Liao P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostreft A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Plattman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Stupski M.P., Smith T.,
RA Spier E., Spradlin A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
EMBL: AE003716; AAF53402.1;
F1Ybase: FB9n0038492; CG4090.
InterPro: IPR002557; Chitin_binding.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF01607; Chitin_bind_2; 10.
DR SMART: SM00494; ChIBD; 11.
DR PROSITE: PS00022; EGF_1; 1.
```



```

QY 534 LAPTTKEPTSTSDKPAETPKGAPTTPKPA-----PTTPKEPAPTTPKGAPTTL 587
Db 432 --PIPPGKFWTTPPLAPAPPEBK-TVVILPPGPSPSEKPNPAPPEPEPKSSPALPP 488
QY 588 KEPAETTP---KKPAKELAPTTKGP--TSTTSKPAETTPKET---APTTPKEPAPTTP 640
Db 489 APAPSMASAVRPPSPPIPPAPAPAPRASMLALPPAPSPPATRLCPPLPSPAPNSP 548
QY 641 KPAETTPETPPPTTSEVSTPTTKEPTTIHKSP----DESTPELSAETPKA--LENSP 694
Db 549 --PAPAPPTPPKLS--ANPPCPVPPAPNRPAPAPAPAPPELPAPDPPTPPVANSP 604
QY 695 KEGVPTTKTAPATKEWTTAOKKTERDLRTTETTTAAPKMTKETATTEKTESKI 754
Db 605 --PAPPAPAPPSALPFVNPAP-----PPTPAPK-----SRPAL 637
QY 755 TATTTQVSTTTODTTPPKITTLKTTLLAPKVTITTKKITTTEIMNKPEETAKPKDRATN 814
Db 638 PAAPPAPAPPVRAATTP-----PPAPAPAPN 665
QY 815 SKATTPPKQKP---TKAPKKPTSTKKPKTMPPVRKPKPTTTPPKMTSTM-----PELNP 865
Db 666 SMALPPAPDPPIPLLATPPAPAPPLPMPSPAPPLPPAPDPPAPPLITINQPPSPPLAP 725
QY 866 TSRIAEMLOTTTRNQTTPNSKLEVENPKSEDAAGAGGET 905
Db 726 VPGAPLAPLPINGRVPFARKNSLI-----GSSSGDT 756

```

Search completed: April 26, 2002, 16:22:33
 Job time: 538 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:22:33 ; Search time 153.5 Seconds

(without alignments)
965.302 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1212

Perfect score: 5416

Sequence: 1 VKDNKKNRTRKKRPPKPPV.....SPIDVTFRNCCKEKFEEK 1013

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5416	100.0	1404	4	092954
2	5407	99.8	1404	4	095X49
3	2284	42.4	1054	11	09JW99
4	967	17.9	1079	5	09N4S7
5	936.5	17.3	1049	5	0917S1
6	865.5	16.0	1795	5	076894
7	853	15.7	1489	10	096449
8	829.5	15.3	401	6	077765
9	808.5	14.9	1315	10	09SPW0
10	802	14.8	1274	5	020007
11	800.5	14.8	1480	10	09LIR8
12	784	14.5	2187	10	041805
13	776	14.3	2187	11	070670
14	739	13.6	2112	5	09YER9
15	715	13.2	763	2	09XDH2
16	707.5	13.1	555	10	09FPG6
17	671	12.4	1151	11	057580
18	667.5	12.3	4880	11	09JUL1
19	667.5	12.3	5085	11	09JRK6

20	660.5	12.2	6677	5	09N435	09N435 caenorhabd
21	659.5	12.2	1229	5	09A185	09A185 caenorhabd
22	657.5	12.1	2284	5	09VPG1	09VPG1 drosophila
23	653.5	12.1	1514	5	09GUV7	09GUV7 leishmania
24	647	11.9	4833	11	09GYX6	09GYX6 mus muscu
25	647	11.9	5038	11	09GYX7	09GYX7 mus muscu
26	643	11.9	2089	4	014676	014676 homo sapien
27	633	11.7	489	10	041707	041707 vigna ungu
28	632.5	11.7	7962	4	010465	010465 homo sapien
29	632	11.7	761	10	09XQ10	09XQ10 arabidopsi
30	632	11.7	6632	5	017362	017362 caenorhabd
31	629	11.6	3507	5	023587	023587 caenorhabd
32	628.5	11.6	1612	5	09VY02	09VY02 drosophila
33	626.5	11.6	990	13	091803	091803 xenopus lae
34	625	11.5	6642	5	001761	001761 caenorhabd
35	624.5	11.5	2768	5	09VC00	09VC00 drosophila
36	622.5	11.5	839	2	09RX57	09RX57 delnoccoc
37	622	11.5	3570	4	099552	099552 homo sapien
38	620.5	11.5	971	5	09XVS4	09XVS4 caenorhabd
39	615	11.4	2344	5	09N3Y8	09N3Y8 caenorhabd
40	607.5	11.2	801	5	023635	023635 caenorhabd
41	607.5	11.2	924	12	099307	099307 epstein-bar
42	605	11.2	379	5	027929	027929 drosophila
43	600.5	11.1	1893	5	09NKC9	09NKC9 drosophila
44	598.5	11.1	409	10	09SBM1	09SBM1 volvox cart
45	577	10.7	956	10	09LJ64	09LJ64 arabidopsi

ALIGNMENTS

RESULT	ID	1	PRELIMINARY:	PRT:	1404 AA.
092954	092954	01-FEB-1997 (TREMBLrel. 02, Created)			
AC	092954	01-FEB-1997 (TREMBLrel. 02, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)				
DE	MEGAKARYOCYTE STIMULATING FACTOR.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,				
RA	Kellerer K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,				
RA	Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,				
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,				
RL	Blood 78:279-279(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Weidberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,				
RA	Fitzgerald M., Scaltreto J., Kellerer K., Preissner K., Kriz R.,				
RA	Jacobs K., Turner K.,				
RL	(In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,				
RL	Mosher D.F. (eds.);				
RL	Biology of vitronectins and their receptors., pp.45-52,				
RL	Elsevier Science Publishers B.V. (1993).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,				
RA	Kellerer K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,				
RA	Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,				
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,				
RL	Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.				
RL	EMBL: U70136; AAB09089.1;				
DR	InterPro: IPR000585; Hemopexin.				
DR	InterPro: IPR001212; Somatomedin_B.				
DR	InterPro: IPR002400; GF_CYS_knot.				
DR	Pfam: PF00045; hemopexin_2.				
DR	Pfam: PF01033; Somatomedin_B; 2.				
DR	PRINTS: PR00438; GFCYS_KNOT.				

```

DR PRINTS: PRO0022; SOMATOMEDINB.
DR PROSITE: PS00024; HEMOPEPTIN; UNKNOWN_1.
DR PROSITE: PS00524; SOMATOMEDIN_B; 2.
DR SMART: SM00120; HX; 2.
DR SMART: SM00201; SO; 2.
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35FAF6 CRC64

```

Query Match	100.0%;	Score 5416;	DB 4;	Length 1404;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1013;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

Oy	1	VKKKKKKKKKKPPKPPVYDEGSGJGNDGRVTTPTDSTJQHNKSTISPKITTAIPIN	60
Db	200	VKKKKKKKKKKPPKPPVYDEGSGJGNDGRVTTPTDSTJQHNKSTISPKITTAIPIN	259
Oy	61	PPPSLPPNSDTSKETSJLVNKETVTEKETTNNKQSTDGNEKTTSAKETQSIEKTSAK	120
Db	260	PPPSLPPNSDTSKETSJLVNKETVTEKETTNNKQSTDGNEKTTSAKETQSIEKTSAK	319
Oy	121	DLAPPSKVLAKPPPKATTTTGGPALTTPKEBPPTPKBPASTPKBPPTTIKSAPTTTPK	180
Db	320	DLAPPSKVLAKPPPKATTTTGGPALTTPKEBPPTPKBPASTPKBPPTTIKSAPTTTPK	379
Oy	181	EPAPPTTSSAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKSAPTTTKEBPAPT	240
Db	380	EPAPPTTSSAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKSAPTTTKEBPAPT	439
Oy	241	PKKPAPTTKEBPAPTTTKEPTPTTTPKEBPAPTTTKEBPAPTTTKEBPAPTTAKKAPATTTPKBP	300
Db	440	PKKPAPTTKEBPAPTTTKEPTPTTTPKEBPAPTTTKEBPAPTTTKEBPAPTTAKKAPATTTPKBP	499
Oy	301	PTTTPKEBPAPTTTKEBPSTTPKEBPAPTTTKSAPTTTKEBPAPTTTKSAPTTTKEBPSTTTKE	360
Db	500	PTTTPKEBPAPTTTKEBPSTTPTTPKEBPAPTTTKSAPTTTKEBPAPTTTKSAPTTTKEBPSTTTKE	559
Oy	361	PAPPTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKKAPATAPKEBPAPTTTKEBPAT	420
Db	560	PAPPTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPATAPKEBPAPTTTKEBPAT	619
Oy	421	TPKKLTPTTTPKLLAPTTPEKAPATTPELAPTTPEBPPTTPEBPAPTTTPEBPAPTTTPEBPAPTTTKE	480
Db	620	TPKKLTPTTTPKLLAPTTPEKAPATTPELAPTTPEBPPTTPEBPAPTTTPEBPAPTTTPEBPAPTTTKE	679
Oy	481	PAPPTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKE	540
Db	680	PAPPTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKE	739
Oy	541	PEPTSTTSOKPAPTTPKGTAPATTPEKAPATTPEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKE	600
Db	740	PEPTSTTSOKPAPTTPKGTAPATTPEKAPATTPEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKE	799
Oy	601	KELAPTTTKEBPSTTSOKPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKE	660
Db	800	KELAPTTTKEBPSTTSOKPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKEBPAPTTTKE	859
Oy	661	PTTTPKEPTTIHKSPEBPSTPELSAEPPTKALENSKEBPVTTTPAPATPEMPTTAKKOKT	720
Db	860	PTTTPKEPTTIHKSPEBPSTPELSAEPPTKALENSKEBPVTTTPAPATPEMPTTAKKOKT	919
Oy	721	TERDLATTPPETTTAPKATTEATTEKTESKITTATTOVSTTQDPTTPKITTLLKT	780
Db	920	TERDLATTPPETTTAPKATTEATTEKTESKITTATTOVSTTQDPTTPKITTLLKT	979
Oy	781	TLAPKVVTTTKTITTTTELIMNKPEETAKPKDRAINSKATTPKPKPKPKAKKPKPTSTIKKPKPT	840
Db	980	TLAPKVVTTTKTITTTTELIMNKPEETAKPKDRAINSKATTPKPKPKPKAKKPKPTSTIKKPKPT	1039
Oy	841	MPRVKRPKPTTPPKKMTSIMPENLNPSTSIAPAMQTTTRNQPNSKIVENPKSEDAG	900
Db	1040	MPRVKRPKPTTPPKKMTSIMPENLNPSTSIAPAMQTTTRNQPNSKIVENPKSEDAG	1099
Oy	901	ABCEBPPHMLLRPHVMEPVTEDMDYLPFRVNOGIIINPMLSDETNICNGKPYDGLTTLLRN	960

	110	115
D6	AEGETPHALLRPHEMEVEVPDDIDYLPFPNGOIIINPMISDETNICNCKRPVDGLTILRN	115
Qy	961 GLTVAFRCHRYFWMLSPFSPSPARRILEVWGIDSPIDITYFRNCNCGKTFEPK	1013
D6	1160 GLTVAFRCHRYFWMLSPFSPSPARRILEVWGIDSPIDITYFRNCNCGKTFEPK	1212

RESULT	2	
Q9BX49		
ID	Q9BX49	
	PRELIMINARY;	
		PRT: 1404 AA.

RA 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR).
 GN BG174L6.2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 OX [1]
 RN RP SEQUENCE FROM N.A.
 RA Wray P.;
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL133553; CCAC36090.1; --
 SQ SEQUENCE 1404 AA: 151076 MW: 782A11746B3FDEE5 CRC64;

Query Match	99.8%;	Score 5407;	DB 4;	length 1404;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 1011; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

[illegible]

```

QY 601 KELAATTTGGPTSTSDKAPATTPEKETAATTPKEBAPATTPKKAPATTPEPTTSEVST 660
DB 800 KELAATTTGGPTSTSDKAPATTPEKETAATTPKEBAPATTPKKAPATTPEPTTSEVST 859
QY 661 PTTTKEPTTIHKSDESPSTBELSAEPTPKALENSPKPEGVPTTKPAATKPEMTTAKDKT 720
DB 860 PTTTKEPTTIHKSDESPSTBELSAEPTPKALENSPKPEGVPTTKPAATKPEMTTAKDKT 919
QY 721 TERDLRTPEPTTAAPKMTKETATTTTEKTTESKATATTQVSTTQOTTEFFKTTTLTKT 780
DB 920 TERDLRTPEPTTAAPKMTKETATTTTEKTTESKATATTQVSTTQOTTEFFKTTTLTKT 979
QY 781 TLAAPVTTTAKTTTTELTINMKPEETAKPKDRATNSKATTPKOKPTKPKKPTSTKTKPT 840
DB 980 TLAAPVTTTAKTTTTELTINMKPEETAKPKDRATNSKATTPKOKPTKPKKPTSTKTKPT 1039
QY 841 MPVKKPKTTTPPKMTSTMBELNPTSRILAEAMLOTTTRPQOTNSKLVEVNPKSEADAG 900
DB 1040 MPVKKPKTTTPPKMTSTMBELNPTSRILAEAMLOTTTRPQOTNSKLVEVNPKSEADAG 1099
QY 901 AEGETPHMLLRPHVFMPEVTPDMMDYLPRVPMQGIINPMLSDEFINICNGKRVGDLTLRN 960
DB 1100 AEGETPHMLLRPHVFMPEVTPDMMDYLPRVPMQGIINPMLSDEFINICNGKRVGDLTLRN 1159
QY 961 GTIAFRGHYFMWLSFSPSPSPARITEWVGISPIDTFTRCNCEGKTEFFK 1013
DB 1160 GTIAFRGHYFMWLSFSPSPSPARITEWVGISPIDTFTRCNCEGKTEFFK 1212

RESULT 3
Q9JM99
ID Q9JM99 PRELIMINARY; PRT: 1054 AA.
AC Q9JM99:
DT 01-OCT-2000 (TIEBMLrel. 15, Created)
DT 01-OCT-2000 (TIEBMLrel. 15, Last sequence update)
DT 01-JUN-2001 (TIEBMLrel. 17, Last annotation update)
DE MRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN PG4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikegawa S., Nakamura Y.:
RT "A novel mouse gene highly homologous to a human gene encoding
RT megakaryocyte stimulating factor precursor and cartilage superficial
RT zone protein."
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB034730; BAA92310.1;
DR MGI: 1891344; Prg4.
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR002965; P_tich_extensn.
DR InterPro: IPR001212; Somatomedin_B.
DR Pfam: PF01033; Somatomedin_B; 2.
DR PRINTS: PRO01217; PRICHEXTENS.
DR SMART: SM00120; HX; 2.
DR SMART: SM00201; SO; 2.
DR PROSITE: PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE: PS00524; SOMATOMEDIN_B; 2.
SO SEQUENCE 1054 AA; 115991 MW; 4FC64BFA24283235 CRC64;

```

Query Match 42.4%; Score 2294; DB 11; Length 1054;
 Best Local Similarity 47.9%; Pred. No. 1.7e-141;
 Matches 491; Conservative 35; Mismatches 130; Indels 368; Gaps 20;

```

QY 1 VADNNKNTTKKPTPKPVVVDAGSGLDNGDFKVT--TPDSTQOHNVSTSPKTTAKP 58
DB 195 VADNNKNTTKKPTPKPVVVDAGSGLDNGDFKVT--TPDSTQOHNVSTSPKTTAKP 254

```

```

QY 59 INPRSLRPNSSDKSTSLTVNKETVETKETTNTNKTSTODGKEXTTSAKETS 118
DB 255 VTKPSLAPNSETSKASLANSKETVETKETTATNKQSSA-SKKTTSKETS 313
QY 119 AKDLAPSKVLAKPTKRAETTTGPAITTPKEPTPTTPKBPASTTPKEPTTIKSAPTT 178
DB 314 KDVV-----EPTSTPK-----NSAPTT 331
QY 179 PKBPATTTKAPPTPKBPATTTKBPATTPKBPATTTKBPATTTKBPATTTKBPATTT 238
DB 332 TKRPV-TTTESKFLP-----LQEPPTAKKEPTTKKBPATTTKBPATTT 376
QY 239 TTKKAPATTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 298
DB 377 TTKKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 436
QY 299 PAPTTPKBPATTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 358
DB 437 PAPTTPKBPATTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 474
QY 359 KEBAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 418
DB 475 KEBAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 512
QY 419 PTTPKKLTTTEKLAATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 478
DB 513 -----TP 514
QY 479 KEBAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 538
DB 515 KEBAPTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 549
QY 539 TKEPTSTSDKAPATTPKGATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 598
DB 550 -----EPTTPKEPVPTTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTP 581
QY 599 AKELAPTTTGGPTSTSDKAPATTPEKETAATTPKEBAPATTPEKETAATTPKEBAPATTPEKETAATTP 658
DB 582 EP-----TTKKEPTTIHKSDESPSTBELSAEPTPKALENSPKPEGVPTTKPAATKPEMTTAKD 617
QY 659 STPTTKEPTTIHKSDESPSTBELSAEPTPKALENSPKPEGVPTTKPAATKPEMTTAKD 718
DB 618 ----- 617
QY 719 KTERDLRTPEPTTAAPKMTKETATTTTEKTTESKATATTQVSTTQOTTEFFKTTTLTK 778
DB 618 -----TSEKT-----TTLK 626
QY 779 TTTLAAPVTTTAKTTTTELTINMKPEETAKPKDRATNSKAT-----TKRP-OKPTKAP 829
DB 627 ATTLAPKVTAAPAE-----EIONKPEETTPASEDDSKITLKPKPKAPATPKKPTKAP 681
QY 830 KKPSTSKKPKTPKPVVVDAGSGLDNGDFKVT--TPDSTQOHNVSTSPKTTAKP 889
DB 682 KKPSTSKKPKTPKPVVVDAGSGLDNGDFKVT--TPDSTQOHNVSTSPKTTAKP 949
QY 890 EVNPKSEADAGSEPTPHMLLRPHVFMPEVTPDMMDYLPRVPMQGIINPMLSDEFINICNG 949
DB 739 EVNPKSEADAGSEPTPHMLLRPHVFMPEVTPDMMDYLPRVPMQGIINPMLSDEFINICNG 997
QY 950 KPVVGLTTLNGLVAVRGHYFMWLSFSPSPSPARITEWVGISPIDTFTRCNCEGK 1009
DB 798 KPVVGLTTLNGLVAVRGHYFMWLSFSPSPSPARITEWVGISPIDTFTRCNCEGK 1057
QY 1010 FFFK 1013
DB 858 FFFK 861

```

RESULT 4
 ID Q9NA57 PRELIMINARY; PRT: 1079 AA.
 AC Q9NA57;

DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE Y51B11A.1 PROTEIN.
 GN Y51B11A.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peleiderinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RX MEDLINE=96069613; PubMed=9851916;
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA None.
 RT "Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium."
 RL Science 282:2012-2018(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Cotton M.;
 RT "The sequence of C. elegans cosmid Y51B11A.";
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL: ACC06797; AAF60743.1; -;
 DR InterPro: IPR002965; P-rich_extensn.
 DR PRINTS: PR01217; PRICHEXTENS.
 SQ SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;

Query Match 17.9%; Score 967; DB 5; Length 1079;

Best Local Similarity 32.3%; Pred. No. 3.2e-55;
 Matches 294; Conservative 114; Mismatches 377; Indels 124; Gaps 33;

QY 35 TTPDSTGTHNKVSTSP-KITTAAPINRPSLPPNSDTS--KEISLVNKETVEIKET 91
 DB 102 TAPETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAPETTSSTPSSS 161
 QY 92 TTNQSTGDKKTKTSAKQSIKTSKDLAPTSKYLAKPFAETTKGP-ALTTPK 149
 DB 162 TTPVQTT-----TTTAPETTSSTPSSS-----TSPVQTTTAPETTSSTPSS 211
 QY 150 EPTPTPKPAATTPKEPTPTTIKSAPTTPKEPAATTTKSAPTTPKEPAATTTKEPA 209
 DB 212 OTTTTAPETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAPETTSST 268
 QY 210 PREPAATTPKEPAATTTKSAPTTPKEPAATTPKEPAATTTKEPAATTTKEPA 269
 DB 269 STEP-PSSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTPV 326
 QY 270 PTTKPAATTPKEPAATTPKAPATTPKEPAATTPKEPAATTPKEPAATTPKEPA 315
 DB 327 OTTTTAPETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAPETTSST 386
 QY 316 SEPTTPKEPAATTTKSAPTTPKEPAATTTKSAPTTPKEPAATTTKEPAATTTKEPA 372
 DB 387 PSSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTPVQTTT 446
 QY 373 PTTKPAATTPKEPAATTPKAPATTPKEPAATTPKEPAATTPKEPAATTPKEPA 419
 DB 447 T-----TTPKPAATTPKEPAATTPKAPATTPKEPAATTPKEPAATTPKEPA 502
 QY 420 ----TTPKPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTP 468
 DB 503 PSSSTTPVQTTT-----APETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTPVQ 558
 QY 469 TTKAAPATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPAATTPKEPA 526
 DB 559 TTTTAPETTSSTPSSSTTPVQTTTAPETTSSTPSSS-TTPVQTTTAPETTSSTP 617

QY 527 KKPAKELAPTTKEPTSTSDKP-APTTPKAPATTPKEPAATTPKEPA--APTTPKGA 583
 DB 618 PSSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAP 674
 QY 584 PTTKPAATTPKPAKELAPTTKGPSTSDKP-APTTPKAPATTPKEPAATTPK 642
 DB 675 TTTAPETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAPETTSSTP 732
 QY 643 PAPTPETTPPTTSVSTPTTPKPTTIKHSPPDESTPELSAEP-----PKLENSPKPE 699
 DB 733 --PSSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTPVQTT 786
 QY 700 PTTTPKAPATTPKEMTTAKDKTTERDLRT-----PETTTAPAKMKKEVATTTKETS 755
 DB 787 OTTTT-----TAPETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAPET 843
 QY 756 ATTQVSTTTTQDTPPEKITTLKTTTAPKVTTK-----KITTTTINKKPELA-KPK 809
 DB 844 STEPPSSSTPVO-----TTTAPETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTP 894
 QY 810 DRATSKATITTKPKPKPTAPKPKPTSKPKPMVRKPKTTPRKMSTMPKELNPSRI 869
 DB 895 SSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAPETTSSTPSSSTTPVQTTTAPET 946
 QY 870 AEAMLQTTT 878
 DB 947 ----VQTTT 951

RESULT 5
 Q917S1 PRELIMINARY; PRT; 1049 AA.

ID Q917S1
 AC Q917S1
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE CG5228 PROTEIN.
 GN CG5228.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sultun G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Adayani A., An H.-J., Andrews-Flandroch C., Baldwin D.,
 RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockett P., Brockett P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Chertys J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos R., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fiesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Wei M.-H., Ibegyan C.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.P.D., Lai Z.,
 RA Kimmel B.E., Kodra C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster".
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003435; AAC22353.1; -
 DR Flybase: FBgn0030561; CG5228.
 SO SEQUENCE 1049 AA; 107278 MW; 954D629E7619671 CAC64;

Query Match 17.3%; Score 936.5; DB 5; Length 1049;
 Best Local Similarity 34.6%; Pred. No. 3e-53;
 Matches 371; Conservative 45; Mismatches 439; Indels 217; Gaps 54;

QY 2 KDNKKNRKPPKPPVNDGAGSLNDGDFKVTPTDTS--TTQHNKVTSPKITTAKP 58
 DB 30 KPNKIGTTAKPTLKP-----TSGTSAKPTTLKPTEGSTSAKPTTLKP 71
 QY 59 -----INPRESLPSNSTSKETSL-----TVNKETTVETKETTNNKQT--STDGKEKPT 106
 DB 72 TEGTTAKPTTLKPTEGTAKPTTLKPTEGSTSAKPTTLKPTEGTAKPTTLKPTEG-----T 127
 QY 107 SAKET--QSIETKSATK--DLAPTSKVLAKPT-----PRAETTTGCPALITPKEPTPT 155
 DB 128 SAKPTTLKPTEGTAKPTTLKPTEGSTSAKPTTLKPTEGSTSAKPTTLKPTEGSTSAKPTTLK 187
 QY 156 PKE-----PASTPKPEPT--PTTISAPPTKEPAPTTTSSAPPTPKPAPPTTEPAP 207
 DB 188 PTEGSTSAKPTTLKPTEGSTSAKPTTLK--PTEGTAKPTTLK--PTGCTSAKPTTLKPTEG 243
 QY 208 TTPKEPAPPTTTPKEPAPTTTTSAPTTTKE--PAPTTPKKAPPTTKEPAPTTTKEPT--P 262
 DB 244 TTAK-----PTTLKTEGTITAKPTTLNTEGSTSAKPTTLKPTEGTITAKPTTLNTEGSTSAK 300
 QY 263 TTPKEPAPTTTKEPAPTTTKEPAPTTTAPKAPPTTTPKEPAPTTTKE-----EPAPTTTKEPSP 317
 DB 301 TTTLKPTEGTITAKPTTLKPTEGSTSAKPTTLKPTEGTITAKPTTLKPTEGSTSAKPTTLKP 360
 QY 318 TTPKEPAPTT-----TKSAPTTTKEPAPTTTAKSA--PTTPKEPSPPTTTPKEPAPTTTKE 368
 DB 361 TTAK-----PTTLKPTEGSTSAKPTTLKPTEGTITAKPTTLKPTEGSTSAKPTTLKPTEGTITAK 416
 QY 369 PAPTTPKAPPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 423
 DB 417 --PTTLKPTEGTITAK--PTTLK--PTEGTITAKPTTLKPTEGSTSAKPTTLKPTEGTITAKPT 470
 QY 424 KLTPT-----TPEKLAPTTPEKAPPTTPEELAPT--TPEEPPTTPEEPAPTTTPEAPAPN 476
 DB 471 TLPKEGTITAKPTTLKPTEGSTSAKPTTLKPTEGTITAKPTTLKPTEGTITAKPTTLKPTEG 530
 QY 477 TTPKEPAPTTTTPKEPAPTTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 526
 DB 531 KCTTAKPTTLK--PTEGTSAKPTTLKPTEGTITAKPTTLKPTEGSTSAKPTTLKPTEGTITAK 587
 QY 527 K-----KPA-----PKELAPT-----TTKEPTSTTSDDKA-----PTTPKGT 558
 DB 568 KPTTLKPTEGSTSAOPTTLKPTEGSTSAOPTTLKPTEGTITAKPTTLKPTEGSTSAOPTTLKP 647
 QY 559 APTTPKEPAPTTTTPKEPAPTTTTPKGTAPTTTLKPAPTTTKE-----KAPKELAPTTTTPKGTST 614
 DB 648 EGTITAK-----PTTLK--PTEGTSAKPTTLKPTEGTITAKPTTLKPTEGTITAKPTTLNTEG 701
 QY 615 TSDKAPPTTPE--TAPTTPKEPAPTTTTPKKA--PTTPEPTPTTSEVSPPT--TTKEPT 668
 DB 702 TSKAPTTTLKPTEGTITAKPTTLKPTEGSTSAKPTTLKPTEGTITAKPTTLKPTEGTITAKPT 761

QY 669 TIKSPDESSTELSAEPTP-----KALENSPKERGV--PTTKT---PAATK 709
 DB 762 TL--KPLEGT---SAKPTTLKPTEGTITAKPTTLKPTEGSTSAOPTTLKPTEGSTSAOPTTLK 816
 QY 710 PENTTTAKDKTTERDIRTT-----PPTTAAPKMKETAPTT-----TEKTES 752
 DB 817 PTEGTITAKPTTLKPTEGSTSAOPTTLKPTEGTITAKPTTLKPTEGSTSAKPTTLKPTEGTITAK 876
 QY 753 KITATTTTQVSTTTTQDTPEFKITTLKTTLAKKVITTK--TITTEIMNKPETAKPKD 810
 DB 877 PTLTKTEGSTSAKPTTLKPTEGTITAKPTTLKPTEGTITAKPTTLKPTEGSTSAKPTTLKPTE 936
 QY 811 RATNSKATTPKOKPTKAPK--KPT--STKKPTMPRVKPKKPTPT--PRKMTSTP- 861
 DB 937 -GTTAKPTTLKPTEGSTSAKPTTLKPTEGTITAKPTTLKPTEGSTSAKPTTLKPTEGSTSAOPT 995
 QY 862 ELNPTSRIAEAMLOTTTTRNQTNSKLVENPKSDAGAGC-ETPHMLLRP 912
 DB 996 TLPKPTERTSAQ--PTTLKPTEGTITAKPTTLKPTEGTITAKPTTLKPTEGTITAKPTTLKP 1045

RESULT 6
 ID 076894
 AC 076894 PRELIMINARY; PRT; 1795 AA.
 DT 01-NOV-1998 (TRENBLREL. 08, Created)
 DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
 DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
 DE EG:5667.1 PROTEIN.
 GN EG:5667.1 OR CG14796.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
 RA Burkitt K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Chin M.A., Chui W., Cleveland L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Harits N.L., Harvey D., Heiman T.J., Gu Z., Guan P., Harris M.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
 RA Jalili M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mouton S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cadiou E., Dreaux S., Lelaure V., Motlier S., Gallibert F.,
 RT "Sequencing the distal X chromosome of *Drosophila melanogaster*,"
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Benos P.,
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AE003421; AAF5644.1; -
 DR EMBL: AL031028; CAI19845.2; -
 DR FlyBase: FBgn0023390; EG:5667.1.
 DR InterPro: IPR002557; Chitin_binding.
 DR InterPro: IPR002965; P-rich_extensn.
 DR Pfam: PF01607; Chitin_bind_2; 2.
 DR PRINTS: PRO1217; PRICHEXTENS.
 DR SMART: SM00494; CnCBD2; 2.
 DR SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match 16.0%; Score 865.5; DB 5; Length 1795;
 Best Local Similarity 28.5%; Pred. No. 2e-48;
 Matches 276; Conservative 129; Mismatches 447; Indels 115; Gaps 28;

QY 6 KNRTKKKPPPPVVDAGSGLDNGDFKVTTPPTSTTQNNKSTSKITAPINRPSL 65
 DB 431 KNDVTEAPEIKSPV-----KGHLSENIVLPEPTTT--TTTTKVVVLCPPLISPDPT 483
 QY 66 PPSSDT-----SKETSLVNNKETVEKETTTNNKQSTIDGKERKTSAKETOSI 114
 DB 484 PKDSTTAATKSTPKKISSIOHSTTAKTTTKRPTTVIEKISSANERKRTVAVTTTQK 543
 QY 115 EKTSANDLAPTSKVLAKTPKAEATTGKPALETTKGPALTTPKEPTTPKEPASTPKETPTIIS 174
 DB 544 RSTTHNTSPDKTTTISTLSKPTTTPSTTTPSTTTPST-----TPSTTTPST--- 594
 QY 175 APTTPKEPAPTTKSAPTTPKEPAPTTK--EPAPTPKEPAPTTTPKEPAPTTTTSAPTTP 233
 DB 595 --TPSTTTSSTTTVAVSHRPTTSOKTTTASTTK-----TTSPTKTKTTDTPDPT 648
 QY 234 KEPAATTPKKRAPTPKEPAPTTKEPTTPKEPAPTTKEPAPTTKEPAPTPKAP 293
 DB 649 TSKISTTQKTTTTHKFTAATSTTEKPTTTEKTSVSTTTKSTESSKPTISSTGKPT 708
 QY 294 TTPKEPAPTTTPKEPAPTTTEPSPPTTPKEPAPTTTKSAPTTPKEPAPTTTTSAPTTP 353
 DB 709 TTPKSTRTPTTTPKVTTPQITTTTPPLRSTSTETSTOPTTTPPOTTTTTLVLVPTKS 768
 QY 354 SPTTPKEPAPTTTPKEPAPTPK-----KPAATTPKEPAP-----TPKE 392
 DB 769 TTTTTEKPIISSKPTTTTQKTSTAPNTKVAITQKETTPROSTSTTITFKTTTNN 828
 QY 393 PAPTPKAPAPPAKEPAPTPK--ETAPTPPKKLTTPTEKLAATTEPKP-----AP 443
 DB 829 PEPTSTKPISTTPKPTTTPKSTSVASSTETKTISPKPTTSTKSTENPTTNSVKTAL 888
 QY 444 TTPTELAATTEPEPTTPKEPAPTTPKAAPTTPKEPAPTTKEPAPTT--TPKEPAPTT 501
 DB 889 TTSSTORASTISEPTKIT--QNTTTTPKPTTLKSTOATSTQKSVITITTKATESS 947
 QY 502 PKETAPTPPKGAPTTTLKEPAPTTPKAPKELAPTTTKEPT--STSDKAPATTPKGA 559
 DB 948 PLTTTISTEEPNTPKPLRTTTPPTTSVATRTITTTTISSESTSTQKKSSTTPST 1007
 QY 560 PTPTEPAPTTTPKEPAPTTPKGAPTTTLKEPAPTTPKAPKELAP--TTTGGPTSTSDK 618
 DB 1008 RTPPKVTIVIVSTONPPTTTSKTSITVTI-----TTP--NPSTQKPTTTTTPQPSITAST 1061

QY 619 PAPTPKEPAPTTTPKEPAPTTTPKRAPTPTEPTPPTSPSTPT--TKE----- 666
 DB 1062 TSGTTRIPPTTTPNPNSTSTDLTTVTRPCDDPSTSKNTACTOELQVNLLEQS 1121
 QY 667 -----PTTIKSPDESSTPELSAEPFPKALENSKREGEVPTTKPATPKEMTTTKD 718
 DB 1122 POKOEPTTHRTHTALGSRNLGCGEVDVMDAPSSAEASGQATAKAPMTSTLAA 1181
 QY 719 KTER--DIRTPETTPAAPPKMTKTATTEKTESKITATTTQVSTTQDTPPKIT 775
 DB 1182 HLQKLFHIIISWTPPSRENAP--TORPSSOPSSQSR--GVTTAOWARHNLATSKPFIH 1238
 QY 776 TLKTT-----TLAKVTYTTKTTTTEIMNKPEETAKPKDRATNSKATTPKPKPA--PK 830
 DB 1239 SLRSTIQQLASTOKRSIPKTLVTHNTKPEPD--SEYDSEISEQYTDDBNVLKTOPR 1297
 QY 831 KPTSTKKKPTMPVYKREKTPPTPKKMTSTWPELNPSTSLAEAMLQTTTPRNPQNSKIVE 890
 DB 1298 AMSSTVAAVLPAVPSTTTREROK--TSSSP-----KATSTTTPQIETTGDLEY 1350
 QY 891 VNPKSED 897
 DB 1351 DSGSSD 1357

RESULT 7
 096449 PRELIMINARY; PRT; 1489 AA.
 AC 096449;
 DT 01-MAY-1999 (Tremblere). 10, Created)
 DT 01-MAY-1999 (Tremblere). 10, Last sequence update)
 DT 01-MAY-1999 (Tremblere). 10, Last annotation update)
 DE CARGO GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
 OS Ca90.
 ON Phytophthora infestans (Potato late blight fungus).
 OC Eukaryota; Stramenopiles; Oomycetes; Pythiales; Pythiaceae;
 OC Phytophthora.
 OX NCBI_TaxID=4787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RACE 1-11;
 RA Goehard B.;
 RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF061185; AAC72308.1;
 DR SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;

Query Match 15.7%; Score 853; DB 10; Length 1489;
 Best Local Similarity 32.3%; Pred. No. 1.1e-47;
 Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;

QY 9 TRKKPTPKPPVVDAGSGLDNGDFKVTTPDSTT-----QHNKVSISPKITTAKPINR 62
 DB 338 TPAPTEKPYDEETTYVEESTYAPTKSEINAPTERMHVAIEKPCDTEVTMAPTEET 397
 QY 63 PSLP-----PNSDT-----SKETSLVNNKETVEKETTT--TNKQISTD 100
 DB 398 TYAPTEETTYAPTEETTYAPTEETPEPEETTYAPTEETTYAPTEETTYAP 457
 QY 101 GKEKTSANETOSIEKTSKADLAPTSKVLAKTPKAEATTGKPALETTKGPALTTPKEPTTPKPA 160
 DB 458 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 514
 QY 161 STTPKEPTPTTIKSAPTTPKEPAPTT--TKSAPTTPKEPAPTT--TPKEPAPTTTPKEPAPTT 217
 DB 515 EETPYEPTTEET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 573
 QY 218 TKEPAPTTTTSAPT-----TPKEPAPTTTPKRAPTPTPKEPAPTTTPKAPPTTTPKE 267
 DB 574 TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 632
 QY 268 P-----AP--TPKEPAPTTTPKEPAPTTTPKRAPTPTPTPKEPAPTTTP 304


```

Db 412 PTCGPPSSSVGKPPSVG-----KPAAPMPPHPDPVSP-----EPL-PE 455
QY 63 PSLP-----NSDTSKSLTVNKETTETKTTTNTKOTSDGKEKTTSAETQ 112
Db 456 PSVPAPAPMPPMLRSPADEIYPT-----PPVPAKSPPGTSPASRGAPLQAOPPAAS 511
QY 113 SIEKTSADLAPTSKY-----LAKPTPKAETTKGPAALTTTKEPTTPKE 158
Db 512 SPAPVAVSSPPAPAVLPPPAKTPSPAPVSPAPVSPQOVKSSPPAPVASSP 571
QY 159 PASTPKEPTTTIKSAPTTKEPAATTTKSAPTT-KEPA-----TTTKEPAPTT 210
Db 572 PPMKSP--PPAPKVASPPMLKSSPPAPVSPQPLKSSPPVLMSTSPVKSPPVP 629
QY 211 KEPAATTTKEPAATTTKSAPTTKE-----PAP--TTP-----KKPAATPKKEPAATTKE- 259
Db 630 VASPPPPVKSPPPLAPVSPVVKLPPLAPAGKSTPPEEKPPPTTPVKSPPPEKSL 689
QY 260 PPTTPPKEPAATTTKEPAATTTKEPAATTAKKPAATTTKEPAATTTKEPSTT 319
Db 690 PPTTLTSPPOKPTTPTTSPKPPPSVETLPPPSKSSPEEVSSPPQAFKSSSP-- 747
QY 320 PKEPAATTTKSAPTTTKEPAATTTKSAPTTKEPSTTTKEPAATTTKEPAATTTK 379
Db 748 ---PAPVS--SPPLKSSPPVPSPPESPPPTKSSPPPLAVSSPPVEKTSPPAVSSP 802
QY 380 TTPKEPAATTTKEPAATTTKPAATTAKEPAATTTKEPAATTTKPLTTPTEKLA 439
Db 803 PTPKSSPPPLAVSSPPQVKTSPAPVSSPPPTKSSPPPLAPVSSPPVEKTSPPAV 862
QY 440 KPAATTTPEELAPTTPEPTTPEEPPTTPEEPAPTTPKAANPTTKEPAATTTKEPA 499
Db 863 SSPLPEKSSPPSSVSSPTTVKSSPPAPLSSPMTKSSPPAHVSSPPAEKSSSP 922
QY 500 TTPKEPAATTTKGT-----PTLKEPAATTTKPKPAKELAPTTTKEPTSTSD 548
Db 923 LABISSPPEKSSPPSPVWEKTSPPAVVSSPPPTPKSSPP--APVSSPPVVKSSP 979
QY 549 KPAATTTTTPKGTAPTP-----KEPAATTTKEPAATTTK--GAATTTTKEPAT 593
Db 980 PPAVSSPPTPKPLPPAPVSSPPVVKSSPPPTVSSPPPTKDLPPPTVSSPPT 1039
QY 594 TPKKPAKELAPTTTGTSTSDKAPAT--PKETAPTTK-----EPAATTTKKAAPT 647
Db 1040 XKLPP--APVSSPPVVKSSPPAPVSLPPTTKRPSRIRVSSPPRVVCCPPTL 1096
QY 648 PETPPTTSEVSTPTTKEPTTIHKSPEDESTPELSAEPPTKALENSPKEGVTTTPAA 707
Db 1097 VSSPSPAPKSLPPTPVSSP-----PPEVKSPPPTPVSSPPPAKSSPPTPV 1146
QY 708 TKPMTTAAKDTTERDLRTT-----ETTTAPKMTKETATTTETKTESKITAITTOV 762
Db 1147 SPP-----ELKSSPPAPVSSPPSAKSSPPAPVSLPPEVKSSPPAPIS 1194
QY 763 STTQDTTPEKITTLKTTTAPKVTTKTITTEIMNKEETAKPKDRATNSKATPKP 822
Db 1195 SPPPAKSP-----PPPA-----PMSLPPVVKSSPPAPVSSP 1228
QY 823 QKTPAKPKPTSTKPKTPMRVRKPKTTPPKKATSTMPELNPTSLAEMLQTTTPNQ 882
Db 1229 PPMKSSPPAPVSSPPAP--VKPPLPAPVSSPPAV-----TSAP-- 1271
QY 883 TPNSKLVAVNPKSDAGAGETPHMLLRPHVMEVTPPMDVLPRAVNOGIIINPMISD 942
Db 1272 -----FKKEBOSTA-----PRAEALPPSPNDIILPPTMAN 1302
QY 943 E 943
Db 1303 K 1303

```

RESULT 10
Q20007

```

ID Q20007 PRELIMINARY: PRT: 1274 AA.
AC Q20007;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)
DE COSMID F35A5.
GN F35A5.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RT Nature 368:32-38 (1994).
RL [2]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Leimbach D.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U46675; AAB52641.1;
SQ SEQUENCE 1274 AA: 138065 MW: 07F6BD0292C1799F CRC64;

Query Match 14.8%; Score 802; DB 5; Length 1274;
Best Local Similarity 28.0%; Pred. No. 2e-44;
Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

QY 6 KNRTRK-KP-----TP-----KPVVDENAGSLONGD-----FKYTPPTSTGHNKYSTS 50
Db 274 KNPTRKWKMPWEDPEVEEVEKPPVPEKAPVLKKRDPAPAKADPSPSAARPEPS 333
QY 51 PKTTAKPL-NRPSLPSPVSDTSKESLTVNKETVETKTTTNTKQSTGKEKTTSAK 109
Db 334 SPVPPPTPVKNVKKYKPPWEDDEPAEVEVKPSAPEKTYVLAKKEDEPSSSTPPSDPS 393
QY 110 ETOSIKTSADLAPTSKYLAAPTPKAE-----TTTKGPA-----LTP 148
Db 394 PKAAAVAVNPROSSPKATPLQADPKAQEVPTPVKNPVKKYKPPWEDDEPVEVROP 453
QY 149 KEPTTTP-----KEPASTTKEP-----TP-TTTSAPTTPKAPATTTKSAPTTKEPAP 199
Db 454 EAPAKKTPLYLRKEPAKADTAKFATSKTPTETPEKKDPVAPROSSKKVAAKDSQAAPR 512
QY 200 TTTKEPA-----PTTPKEPAATTTKEPAT-----TTKSAPTT 233
Db 513 TPVKNPVKKRPPWEDDETPADVSKPTDAKKTPLSLAKKOPAPAKESLKPADTKAPAKP 572
QY 234 KEP-----APTTKKAAPTTPKE-----PAPTTPKEPTTTPKEPATTTKEPATTPK-EP 283

```



```

Db 573 RDSPPKVAFTAEKTPVLAKKAPGADSKTEPEKSKPRDPSKAVPAKVPKTEV 632
QY 284 APTAKPKAP-----TTPKEPAPTPPKKAPPTTKEP-----SPTTPKAPAPTTTKASAP 333
Db 633 APAAVAKKPEISKPKDTAKKAPKAPNSVVP--PTPVAKNPKKKKPMWEDDADAKVSLPE 691
QY 334 TTKEPAPTTTKASAPTP--KEPSPTTKEPAPTPPK--EPAPTPPKKAP--TTPKEPAP 388
Db 692 PEKK--TPVLAKKAPTPDSEAAADPVSGSSKDPKLAKKAPKRPDPSMKAVPLKAPK 750
QY 389 TTKEPAPTTTKKAPAPTA--PKKAPPTPKTAPTPKKTLPTR-----EKLAPTPPEKA 442
Db 751 T---EVPAPVAKKPEPVAKSRDPSKAK--AEPSNP--VVPAPVAKNPKKKPMWEDDADA 805
QY 443 PTPPELAPTPPEP-----TTPTEEPAPTPPKAAAPTPKEPAPTPPKKAPPT--PKE 496
Db 806 PAEPVAVPEPEKKTPLAKKTPVKKRDPSPKAVPAKSTKDAPVAVKKEEVSKEKE 865
QY 497 PAPTTPKEPAPTPPKGTAPPTLKEPAPTPPKKAPKELAP--TTEPSTTSDKRAPTP 555
Db 866 PSPKAPNSPVVP-----PTPVKNPVKKW--KPMWEDDEPTEEVKKPSE--PEKTPVLA 918
QY 556 KGTAPTTKEPAPTPPKKAPPTPKGTAPPTLKEPAP-----TPKKAP-----KEL 603
Db 919 K-KEPEKPD--APKVAAPROPSKAPVPE--KEPAKVAAPRDLSPKKAIPDANTQEA 974
QY 604 APTTKGTPS-----TSDKP-----APTTPKET-----APTTPKEPAPTPPKKAP 646
Db 975 PTPVAKNPKKKKPMWEDDDEPAEPVSAPEPEKTPVLAKKAPKAPNDP--SKKAPV 1031
QY 647 TTPPTPTSEVSTPTTKEPTTHKSP-----DESTPELSA-EP--TPVALENSKEPG 698
Db 1032 AAK--PDPRIPEV--PTPVAKNPKKKKPMWEDDDEPSEVSAPEPEKTPVLAKKAPKPA 1089
QY 699 V-----PTTKTAPAT-----KPEMTTAKDKTERDLTTPETT--TAPK 737
Db 1090 TKPDEAAADPVSGPTSKDPKLSKKAVERKPTTDDKDKLSPKKEKAPKEPAP 1149
QY 738 MTKETATTEKTESKITATTTQVSTTTQDTPTEFKITTLKTTTLAPKVTTKTITTE 797
Db 1150 KKKVWDDDDPEPADPTVPAPSKKPTEDPADLG-----GPKTKDPK----- 1193
QY 798 IMNPEETAKKDRATNSKATTPKOKPTKAPKPKPTSTKKKPTMPRRKK----- 848
Db 1194 -LNKKAPEKTEK-----PKPKVSKPEKPTPEPKP-AAFKKMKPMWEDDDEPE 1243
QY 849 ---TTPPKMTSTMP 861
Db 1244 ADFTMAPKPKDTEDP 1259

```

```

RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by niney p1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL: Ap001306; BAB03062.1; -.
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS: PRO1217; PRICHEXTENS.
SQ SEQUENCE 1480 AA; 147153 MW; D1AC0C79P155E732 CRC64;

Query Match 14.88; Score 800.5; DB 10; Length 1480;
Best Local Similarity 29.68; Pident. No. 2,9e-44;
Matches 307; Conservative 54; Mismatches 505; Indels 171; Gaps 45;

QY 10 KKKTPKPVVDEAGSLDNGDKVTPRDSSTQHNKVS-----FKITAKINRPSL 65
Db 117 KPRHPRPVVK-----PHNKRPTKPRHPRKPRKNNHPRPTIKRPPKPSV 167
QY 66 PPNSDTSKETSIVNKEETVETKTTNKKQSTDKREKTSAKETQSIKTSKDLAP 125
Db 168 KPPSPSPK-----PPTTNPPSTPQPRTHKKRPPSPTPPVASPPAPTP 211
QY 126 SKV--LAKPRKAEITTKGPAITTKKPP--TTP--KEBASTPKKPPPTTKASAPTP 179
Db 212 TOMPPATP--PLKSPVAPPTATPPTATPPIIPVAPPTITTPVAPPTIIMPATP 270
QY 180 KEBAPTT-----KSAPTTPKEBAPTT-----KEBAPTPKKAPPTTKE 220
Db 271 PVAPAPITNPPISKRPVTP--PTTTPPIAKPPIATPPISTPPAPVAPPTPI 326
QY 221 PA-----PTTKSAPTP--KEBAPTTKKAP--TTPKEBAPTPKKPPPTKEBA 270
Db 327 PAKPVASIVIT--EPVPIAQPVATPPTATPVAPVATPPIATPPIKSPISSTP 384
QY 271 TTKEPAPTP--KEBAPTAKKAPPTPKKAPPTKEBAPTT-----KEPSPTTKEBA 325
Db 385 VATPTATSPITTPPAKPPVATPPIAKSPITATPAPVAPVATPIEKPPVATPTTP 444
QY 326 TTKSAPTTKEBAPTTTSKAPTPKEP--SPTTKEBAPTPKKAP--TTPKAP 380
Db 445 TAT--PVAKEPVEPTATPPTAKPISPTPISKPVAPVAPVATPTTPPVKPV 501
QY 381 TTKEPAPTPKPP--APTTKKAPATP--KEBAPTTKEBAPTT-----PKLTP--T 428
Db 502 TPPLAPVAKPPVATPPIATPPIATPPIAKSPVATPPIATPPIAKPPVATPTPT 561
QY 429 TPEKLAFTPEKPADTPEELAPTT--PEEPPTPEEPAPTPPKAAAPNT--PREAP 484
Db 563 TPPTATPVAKPPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPI 621
QY 485 TPEKAPPTPKKAPPTTKEKAP--TTPKGTAPTT--LKEBAPTPKKAPKELAPTT 539
Db 622 TPPIATPPIAKPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPI 681
QY 540 KEPSSTSDKAPPTPKGTAP--TTPKEBAPTPKKAPPTPKGTAPTTKEBAPTTKE 598
Db 682 KPVAT--PPTAPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPI 736
QY 599 APKELAPTTGSPST--TSDKAPPTPKKAP--TKEBAPTPKKAPPTT-----E 649
Db 737 MPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPI 796
QY 650 TTPPTSEVSTPTTKEPTTH--KSPDESPELSAEPKALENSKEGCVPTTKPAA 707
Db 797 TPPIAKPPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPI 856
QY 708 TTKEMT--TKAKTTEEDLKTTPETTTAAKMKETATTEKTESKITATTTQVSTTT 766
Db 857 TPIITTPPAKPPVATPPIATP--IAKPPVATPPIATPPIATPPIATPPIATPPI 914
QY 767 QDTPFKITTLKTTTLAPKVTTKKT-----ITTEIMNKKKEPAKDRATNSKATTP 822

```

Dd	915	PAPKPVALPPATPEPAVKPVPVATPPATPIANSPITAPPVV---	PPRTASVATPPPIA	971	
Oy	823	OKPRKAPKPKSTKKPKIMPRVRKPTTPTRKMTSTMELNP-----	TSRLAE---	871	
Dd	972	KPPITTT--PRAT-PPVAMPPIATP----	PRAKPPVATPPPIANPEVEKPVPATPPAPPT	1025	
Oy	872	----AMLOTTTRENQTFNSKL--VEVNPKSSEDAGAGEPHMLLPHVEMPEVTGDMD	924		
Dd	1026	VLPRIAKPPEVESPTAPPTPTATPPVALPP-----	VXKPPVAIPITTKPPV	1070	
Oy	925	VLEPRVNOGIINPMLS	941		
Dd	1071	ATPPVTNPPETAMPPIVT	1087		
RESULT	12				
ID	041805	PRELIMINARY;	PRT; 1188 AA.		
AC	041805				
DT	01-NOV-1996	(TREMBLrel, 01, Created)			
DF	01-NOV-1996	(TREMBLrel, 01, Last sequence update)			
DR	01-JUN-2001	(TREMBLrel, 17, Last annotation update)			
OS	EXTENSIN-LIKE PROTEIN PRECURSOR.				
DE	Zea mays (Maize).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;				
OX	Panicoidae; Andropogoneae; Zea.				
NCBI_TaxID=4577;					
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=B73; TISSUE=POLLIN;				
RU	Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;				
RA	Submitted (JUN-1994) to the EMBL/GenBank/DDBJ databases.				
DR	EMBL; Z34465; CAA84230.1; -				
DR	Mendel; 14346; Zeama; 2368; 14346.				
DR	InterPro; IPR003592; LRR_out.				
DR	InterPro; IPR001611; LRR.				
DR	Pfam; PF00560; LRR_3.				
DR	SMART; SM00370; LRR; 4.				
KW	Signal.				
FT	SIGNAL				
SO	SEQUENCE	1	27	POTENTIAL.	
		1188 AA;	120980 MW;	2C77C7F8D7130149 CRC64;	
Query Match		14.5%;	Score 784;	DB 10; Length 1188;	
Best Local Similarity		29.9%;	Pred. No. 2.8e-43;		
Matches 230; Conservative		59;	Mismatches 359;	Indels 122; Gaps 33	
Oy	13	PTPPEPVVDAGSGLDNGDKVTTPDTSTQHNKVSTSPKITTAKDINPRSL----	PPN	68	
Dd	460	PTPISSPAD-----	DYVPPPPVPGKSPATSPSPGVCPAPASjppPSLVKLSPQO	510	
Oy	69	SDTSKETSLIVNKETTVEIKETTTTNKOISTDGKEKTSAKEQSIETKNSAKDLAPTISKY	128		
Dd	511	APVG-----	-SPPPVKITSPPAIG-----	SPSPPPVSIV	541
Oy	129	-----LAKPPKAETTTKGPAITTPKEDPTPTTKEPASTTPKEDPTTIKSAPTTKEPA	183		
Dd	542	SPPPVPKSPPPPAVGSPPPEKSPPPPAVASPPPVKSP--	PPPLIVASPPPVKSP	599	
Oy	184	PTTYSADPTT--KEPATTTKEPAPTTKEPAATTTKEPAATTTKSAPTT--KEPATPT	240		
Dd	600	PPAPVASPPPVKSPPTTPVASPPPAVASPPPMKSPPPPVASPPPEKSPPPPP	659		
Oy	241	PKKRAPTPPKEPAPT-----PKE--	PPPTTKEPAPTTKEPAPTTKEPAATAFKP	291	
Dd	660	PAKSTPPEEPTPTSVKSSSPPPEKSLPPTLLIPSPPOKTPPTGPSPK--	PSSEPK	718	
Oy	292	AATTKEBAPTTPKEBAPTTTKEBPTTPKEBAPTTKSAPTTTKEBAPTTPK	351		
Dd	719	SP--PKEPVSSPPQR-----	KSSPPPAVSSPPPTPVSSPALAVSSPPSKSP-----	768	
Oy	352	EPSPTTKEBAPTTPKEBAPTTPKBPATTTKEBAPTTPKBPATTTTKKAPATPAKBPAP	411		

	DB	PPAPLSSPPAPQYKSSPPEVQVSSP - PPAPKSSPPPLAIVSSPPQVEKTSPPPAVLSSPP	827
Oy	412	TTPKETAP---TTP-----KKLTPTTEPKIAPTTPPEKPAD-----TTPBELAFTTREET	458
Db	828	LAPSSPPHHVVSSPPPVVKSSPPPAVSSPPLTKRASPAAHVHSSSPEVKESTP--PA	885
Oy	459	PTTEEPAPPTPKAAPATPKE--PATPTTKEBPATTTTKEBAPTTPKETATTTKGAPTT	517
Db	886	PTTVISP--PSEDPKSSPPTPVSLPPPITVKSSPPAVVSSP--PMTPKSSPPPVVSSPPT	943
Oy	518	LKE-----PA-----PTTPKK---PAKELAPTTEE--PTSTSDKAPATTGKAATPD	563
Db	944	VKSSPPPAVPSSPATKSSPPPAVNLPREEVKSSPPPIPVSSPPA---PKSSPPAP	1000
Oy	564	-KEPAPTTPKEBAPTTPKGAATTLKEBAPTTPKKAPAKELAPTTTGPISTTSODKAPT	622
Db	1001	MSSPPEPVKSSPPPAVSSPPEVKSSPPPAVSSSP---PPKSSPPPAVSSSPPV	1057
Oy	623	-TPKETATTTKEBAPTTPKKAPATTPETPTTSEVSTPTTKEPTTIHKSPDESTPEL	681
Db	1058	KAPPAPALISSPPPVKSSPPPAVS--SPPPVKSSPPPAVSSPPIKSPPPAPVS	1115
Oy	682	SAPPTPKALENSPKEDPGVTTKTPAATKPEMTTAKDKTTERDRIKTPTT	731
Db	1116	SPPAP-----VKPPSLP-PPAPVSSPPVTPAPPKESOLPPPAES	1158
	RESULT	13	
	P70670	PRELIMINARY;	PRT; 2187 AA.
AC	P70670:		
DT	01-FEB-1997 (TREMBLrel. 02, Created)		
DT	01-FEB-1997 (TREMBLrel. 02, last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)		
DE	NASENT-POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,		
GN	MUSCLE-SPECIFIC FORM GP220).		
NC	NACA.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
RN	NCBI_TaxID=10090;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-96312450; PubMed-8698236;		
RA	Votow W V., StrArnau R.;		
RT	"differential splicing-in of a proline-rich exon converts alphaNAC		
RL	into a muscle-specific transcription factor.";		
EMBL	Genes Dev. 10:1763-1772(1996).		
DR	EMBL; U48364; AAB18734.1; -		
DR	EMBL; U48363; AAB18732.1; -		
DR	MGI:106095; Naca.		
DR	InterPro: IPR002715; NAC.		
DR	InterPro: IPR003037; TS-N.		
DR	Pfam: PF01849; NAC; 1.		
DR	Pfam: PF02094; TS-N; 1.		
SO	SEQUENCE	2187 AA; 220599 MW; 003646AA864DEBFD CRC64;	
	Query Match	14.3%; Score 776; DB 11; Length 2187;	
	Best Local Similarity	27.7%; Pred. No. 1,6e-42;	
	Matches 288; Conservative 122; Mismatches 389; Indels 240; Gaps	49;	
Oy	34	VTPEDTSTTOHNKSTSPTKITAKPINRPSLDPNDSJSE-----TSLTIVKETT	84
Db	806	VQSKRVDPIMSDVIPTSPKTSATAV-PK-----DTSATLSLKSVPAVTSISPRAVY	857
Oy	85	VEIKETTTNKQISTDGKEKTTSAKFQSLEKSANLAITSVILAKPKPAETTTKGPA	144
Db	858	APSNEATVPEIPTSLKNALAATPKETLATSIPTKTSPS----DQKTPKSVSLKGAR	913
Oy	145	LTT-----PK-----PTPTPKPEASTT-----KREPTPTT	171

Db 914 MTSKATTELAASKDVSPSPQFPKEVPLLOHVPTSPKSPVSDTLGALTSPPKGP-PAT 972
 Qy 172 IKSAPTTPE-PAPTTTSAPTTPK-----EPAPTTKEPAPTTKEPAPTTTEP 221
 Db 973 LAETPTPKSPKPAASKTATPATPEGEVTAAPLEIPPCSKAKPATAAPKSSATSSSKR 1032
 Qy 222 APTT--TTSAPT-----TPKEPAPTTPK-PAPTTKEPAPTT 257
 Db 1033 AKRTAVSKSEIPSKGVTAAPLEISLPLKETSATPEGKSSAPKSPKATPAKE---TPP 1089
 Qy 258 KEPTPTTPEPAPTTKEPAPTTKEPAPTTAPKAPTT--KEPAP---TTPEPAPTT 311
 Db 1090 GGVTAAPPEISIPKETEQONATPNEISLAASSOKRSPKTSVKEPPEPGVTAAPLEISAP 1149
 Qy 312 TKEPPTTPK-----EPAPTTTSAPTTTPKEPAPTT--KSAPTTPK-----EPSPPT 357
 Db 1150 OKAPTAAPKQIPTPEDAVTILAGSPLSPKASKTAAPKEAPATPSGVAVSEISDSP 1209
 Qy 358 TKEPAPTTKEPAPTTPKAPTT--KE-----PADTPEPAP-----T 396
 Db 1210 KTSKTAAPKENSATLPPKRSPTAAPKEPTATSEGVTAAPSEISPSPTPAKGVPT 1269
 Qy 397 TTKRPAPAPKEPAPTTKEPAPTTPKLPTPEKLAPTPEK-PAPTTPEELAPTTPE 455
 Db 1270 LTPGAPVALAE-SPASKKVPKTAAPETSTP-----SPOKIPVAGPKASATPPS 1322
 Qy 456 EPTPTT-----PEEPAPTTKAAAPNTPEPAPTTPE-PAPTTKEPAPTTKET 506
 Db 1323 KRTKTAAPKETSAPSEGVTAAPLEIPSPKAKPATAPEPTAPPS--PEGATTAPQIP 1380
 Qy 507 PTTKGTAPTTLKEPAPTT-----KRAPKLAATTTKEPTS-- 544
 Db 1381 PSPKSKKAGSKE-TPTTSPSEGVTAAPLEIPISSKTSKMASPKETLVTTPSKLSQT 1439
 Qy 545 -----TTSKAPAPTTKGTAPTTKEPAPTTPE-PAPTTKGTAPTTLKEPAPTTKPA 599
 Db 1440 VGPKESTLEGATVAPLEIPSPKAKPVTDPKQVPLPPSK-DAPTTLAE-SPSSPK-A 1496
 Qy 600 PKELAPTTKGTPTSTSDKAPPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTPEP-----PPT 654
 Db 1497 PKTAAPSER-VTVTPPEKA-TPKASGTTASKVPAETOVAAVSSRTPTPAVPPV 1554
 Qy 655 TSEVSTTTTK-----EPTTIKSPDESTPELSAETPKALENSKE-PGVPTTKPAA 707
 Db 1555 KNSSSHKTKTTELKEAPATLPPSPKIPSSKAPRT--SAKEFPASSTIR-PVT 1611
 Qy 708 TPEMTTAAKXTENDLRTTPEPTTAAPKMTETATTEKTESKITAATTQVSTTQ 767
 Db 1612 T--SLAQTAPPSLOKAPSTTPEKNTLAAPV---LPVSSKSPAPARASASISPATAAP 1665
 Qy 768 DTPPKITTLKTTTLAPKVTTTKITTTTINMKPEETAKPKDRATNSKATTPKQKTK 827
 Db 1666 QTAPKEATTPSCKAATETPIETASLSEAPKETSE-----TSVAVLMSPP----- 1716
 Qy 828 APKPTSTKPKTPMPVAPKPTTPPKMTSTWPELINTPSRIAEMALQTTTPPNOTPNSK 887
 Db 1717 -PKKASSSKRASTLP-----ATLPLSKEASVLS-----PFAATSSK 1752
 Qy 888 LVEVNPKSEDAAGAGETP 906
 Db 1753 DSHISPV-SDACSTGTTT 1770

RESULT 14
 QVEL9 PRELIMINARY; PRT: 2112 AA.
 AC QVEL9;
 DT 01-MAY-2000 (Tremblrel. 13. Created)
 DT 01-MAY-2000 (Tremblrel. 13. Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17. Last annotation update)
 DE CG4090 PROTEIN.
 GN Drosophila melanogaster (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gockayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zheng Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,
 RA Adair J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busan D.A., Butler H., Brockstein P., Brothier P.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasio P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merulov G., Milshina N.V., Modarri C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon R., Nuskern D.R., Paclob J.M.,
 RA Palazzolo M., Plittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Rehert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spleer E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhu D., Yang S., Yao Q.A.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Glids R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster".
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003716; AAF55402.1;
 DR FLYbase: FBgn0038492; CG4090.
 DR InterPro: IPR002557; Chitin binding.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF01607; Chitin_bind_2; 10.
 DR SMART: SM00494; ChEBD2; 11.
 DR PROSITE: PS00022; EGF_1; 1.
 SQ SEQUENCE 2112 AA; 219547 MW; B91018E5515A036 CRC64;

Query Match 13.6%; Score 739; DB 5; Length 2112;
 Best Local Similarity 27.2%; Pred. No. 4e-40;
 Matches 281; Conservative 98; Mismatches 339; Indels 316; Gaps 44;

Qy 4 NKKNRKKKPPRPVYDEAGSGIDNGDFVYTPDSTTGHNVSPKTTAKPIPNRP 63
 Db 1023 NHPDQIQTKPKCKKVVQSGSSN-----STNSSSSSSSNSGSSSSGSSSS----- 1070
 Qy 64 SLPPNSDTSKETSLTVKETTIV---ETKETTNNKOS-----IDGKEKTSAAETOSI-- 114
 Db 1071 SSSSSSSSSNTGSSSSNSGSSGSSGSSGSSSSSSSSGSSGSSGSSSSSSSSSS 1130
 Qy 115 -----EKTSADLAPTKSVLAKPPKAEYTTGP----- 143
 Db 1131 NNNNGSSSSSSSSSTSKPMPSETCKVNGOIFGRSDCAKFRVCVNDRGGFNMPF 1190
 Qy 144 -----ALTPKPTPTTPKEPASTTPKEPPTTINSAPTT 178

```

Db 1191 SCGPCTVWDACQACNHAAVKECGSIAPPTTSPTTSR-PTTASTSPSDOTSRSPTG 1249
OY 179 PKEPAPPTTKSAPTTKEAPAPTTTKEAPAPTT----- 217
Db 1250 P-----PTTARPTVABPTTSSPTTASSQTSPTVQAPNTDCKCRSEGMADPNNSKRYR 1305
OY 218 -----TKEP-----APT-----TTKSAPTTKEAPAPTTPKKAPAPT 248
Db 1306 CVARNKKGFTSIFPOCAGTVDODLQCNHNFNNCSITESTTTKPPC-EPAINGTTA 1363
OY 249 PKEPAPTTKEAPPTTKEAPAPTTKEAPAPTTKEAPAPTTAKKAPAPTTKEAPAPTTKEAPA 308
Db 1364 TSSSTTTP--PPTTDLPTSTTGLP-PTTTELPLPT-----TTTDLPTTTRLLRP 1413
OY 309 PTTTAKPPSTTKEAPAPTTTSAPTTKEAPAPTTTSAPTTPK-----EPSPPTTKEAP-- 363
Db 1414 TTTTSLPPTTTGLPPTTTTGAQPTTTTSSSTETSTVTTSBSTTQPPSTTTMKPLPAG 1473
OY 364 -----TPPK-----TPPK-----BPAP--TT 373
Db 1474 TECTGEGYADPEDCKRYKCIAGASYKYNFTCPKGTGMNEVOQCDYVENIPRCSKL 1533
OY 374 PKKAPAPTT--KEAPAPTTKE-PAPTTAKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 428
Db 1534 PABETTTTPEESKDPSTPOSTDEPTVTKPI-TKPE-EPSTKPOKPTTQYPEKPT 1591
OY 429 TPEKAPPTTKEAPAPTTPEELAPTTPEEPTTPEEAPAPTTKAAAPNTKEAPAPTTKE 488
Db 1592 TTEB--PEKQOKP-TEYQOKPTTEEPPEKPOKT-TEYQOKPTTEEPPEKPOKT 1647
OY 489 PAPPT-KEAPAPTTKEA-----PTTPK--GAPPTTKEAPAPTTKAPAPAPAPAPAP 540
Db 1648 PTTTEYQOKPTTEEPPTTSLIPYNPPTTSVPGYNPTTPIPVETTTSTPGYK--PTTGG 1704
OY 541 EPTTSDKAPAPTTKGTAPPTTKEAPAPTTKE-PAPTTAKGTAPPTTKEAPAPTTKE-- 597
Db 1705 EPTTTT-LBSTTTDAIOEPTTSKAPPTTTTSSBSTTSGSTTQPPPOQNYNCSSE 1763
OY 598 ---PAPKELA-----PTTGGPTSTSDKAPAPTTKETAAPPTTKEAPAPTTPKKAPAPTT 648
Db 1764 GFFPDPECSRYKCVDAANKGYOYAFKCGKGTWDTSTENCYNADOVSGN----- 1816
OY 649 ETPPTTSEVSTPTTKEPTTIHKSDESPTELSAEPTRALENSKPEPGPTTKTAPAT 708
Db 1817 ---CSSGQTTTPTGTTTEPTTSTSSGK-----ETTSKAPEN-----TTTWA-- 1857
OY 709 KPEMTTAKDKTTERDLRTTPEETTTAAPKMTETATTTTEKTESKITATTTQVSTTTOD 768
Db 1858 -PETTTT-----SSPETTTT--VASETTTTTSGT-----TTTAPDETTPK 1895
OY 769 TTFKITTLKTTLTAPKVTTKTITTTTEIMNKPEETAKKPRDRAITNSAKITPKQOKPTKA 828
Db 1896 PKP-----ETTTTIGEETSTKSPTTTE-----SPASTNTSTA 1928
OY 829 PKKPTSTKKPTMP 842
Db 1929 P-----CPETGP 1935

```

```

RESULT 15
OYXDH2 PRELIMINARY: PRT: 763 AA.
ID OYXDH2:
AC OYXDH2:
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=173;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Espitia C., Lacleite J.L., Mondragon-Palomino M., Martens A.,
RA Zhang Y., Moreno C., Singh M.;
RT "Cloning and characterization of a new member of the PERK family that
RT is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DDAJ databases.
DR EMBL; AF071081; AAd41594.1; -.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR003882; Pistil_extensin.
DR PRINTS; PRO1222; ATROPHIN.
DR PRINTS; PRO1217; PRICHEXTENSIN.
DR PRINTS; PRO1218; PSTIEXTENSIN.
DR SEQUENCE 763 AA; 75034 MW; 39168EC45A5916F8 CRC64;

```

```

Query Match 13.2%; Score 715; DB 2; Length 763;
Best Local Similarity 28.2%; Pred. No. 5,7e-39;
Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

```

```

OY 132 PTPKAPPTTKEAPALTTKEPTTTPKAPASPTTKEAPPTTISAPTTKEAPAPTTKSAP 191
Db 3 PVP-----APALAPLPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 46
OY 192 TTPKEAPAPTTTKEAPAPTTTKEAPAPTTTKEAPAPTTTKEAPAPTTTKEAPAPTTTKE 251
Db 47 PCP--PAP-----PAPPKKSKAPPPVPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 90
OY 252 PAPPTTKEPTT-----KEAPAPTTKEAPAPTTK-EPAPTPK-KAPAPTTKEAPAPTTK 305
Db 91 SRPAPPCPPPPVVIIDPEAPVPAPAPNSPPPPPPAPKAPAPAPV--EVPNSPP 148
OY 306 EP--APTTKEPSPTTKEAPAPTTTSAPTTKEAPAPTTTSAPTTKEPSPTTKEAP 363
Db 149 PPPPPAPALNPAP-----PAPPLANSPLPAPPTPAGT--PAPAPVPVAPAPAKSKPA 201
OY 364 TTPKEAPAPTTKAPAPTTTKEAPAPTTTKEAPAP-----TTTKKAPAPAPAPAPPTTKEAP 419
Db 202 SPPRPAP-----PMPATMEFPPLPVPDPDISKETPAPAPAPIPAPVPIPVPLP 256
OY 420 TTPKTLTPTTPEKLA-----TTPKEAPAPTTPEELAPTTPEEPTTPEEAPPTT-- 469
Db 257 PVPNKKIPAPP--APVAVAVAVAPCPPLPLPLNNHPAPAPAPVGCPLAPLPNSHP 313
OY 470 ---PKAAAPNTKEAPAP----- 483
Db 314 PAPPSAPVPGVPLAPLISGRPVSWKGSFTTLSTPCRCVSCGEVLALNPSRSRSL 373
OY 484 --TTPKEAPAPTTKEAPAPTTKEA-----PTTPKGA--PTTAKAPAPTT-KKAPKE 533
Db 374 TTTTPALPAPIPPLPPLPPLINTAVPPIPLPVTALAPPLPPLAPLISGVPAP-- 431
OY 534 LAPTTTKEPTSTSDKAPAPTTKGTAPTTKEA-----PTTKEAPAPTTKGTAPPTL 587
Db 432 --PIPGKFWTTPPLAPAPPEK-TYVLPBGSCSPSEKPNPAPPEPEKSSPALP 488
OY 588 KEAPAPTT--KKPAPKELAPTTKGP-ISTSDKAPPTPKT--APTTPKEAPAPTT 640
Db 489 APAPSMPSAVRVPSPPIPPAPAPARASMAPAPAPSPATFRLCPPLPSPSPAPNSP 548
OY 641 KKAAPTTPEPTTSPSTSVSTTTPPTTKEPTTIHKS-----DESPELSAEPTRKA--LENSP 694
Db 549 --PAPAPAPPPKLLS--ANPPCPVPAPANNRRPAPAPAPAPAPAPAPAPAPAPAP 604
OY 695 KEPPVPTTTPAPATKEMTTTAKDKTTERDLRTTPEETTTAAPKMTETATTTTKESTKI 754
Db 605 --PAPAPAPAPAPALPVPNPAP-----PPTPAPAPK-----SRPAL 637
OY 755 TATTTQVSTTQOTTPPKITTLTTLTAPKVTTKTITTTTEIMNKPEETAKKPRDRAIT 814
Db 638 PAAPAPAPAPVRAATTP-----PPAPAPAPAPV 665

```

QY 815 SKATTPKQKP---TKAKKPTSTKPKXTMPVRKPKTTPPRKMTSTM-----PELNP 865
| | | | | : | | | | | : | | |
DB 666 SMALPPAPDPPIPLIATPPAPAPPLPMSPPAPPLPRAAPDPAPPLITINOPSPPLAP 725
QY 866 TSRIAEAMTQTTTRPNQTPNSKLVENVPKSEDDAGAEGET 905
| | | | | : | | | | | : | | |
DB 726 VPGAPLAPLPIINGRPVFARKNSLI-----GSSSGDT 756

Search completed: April 26, 2002, 16:22:47
Job time: 552 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:22:02 ; Search time 153.5 Seconds

(without alignments)
922.421 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1167

Perfect score: 5155
Sequence: 1 VKDNKKRRTKKRPPV.....GKPDGLTLRNGTLVAFRC 968

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_17:*
2: SP_Archaea:*
3: SP_Bacteria:*
4: SP_Fungi:*
5: SP_Human:*
6: SP_Invertebrate:*
7: SP_Mammal:*
8: SP_Mhc:*
9: SP_Organelle:*
10: SP_Phage:*
11: SP_Plant:*
12: SP_Rodent:*
13: SP_Virus:*
14: SP_Vertebrate:*
SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5155	100.0	1404	4	Q92954
2	5146	99.8	1404	4	Q9BX49
3	2046	39.7	1054	11	Q9JMG9
4	967	18.8	1079	5	Q9N4S7
5	936.5	18.2	1049	5	Q917S1
6	865.5	16.8	1795	5	Q76894
7	853	16.5	1489	10	Q96449
8	808.5	15.7	1315	10	Q9SPM0
9	802	15.6	1274	5	Q20007
10	800.5	15.5	1480	10	Q9LITE8
11	784	15.2	1188	10	Q41805
12	776	15.1	2187	11	P70670
13	739	14.3	2112	5	Q9VEL9
14	715	13.9	763	2	Q9XDH2
15	707.5	13.7	555	10	Q9FPG6
16	671	13.0	1151	13	Q57580
17	667.5	12.9	4880	11	Q9JLTI
18	667.5	12.9	5085	11	Q9JKE6
19	660.5	12.8	6677	5	Q9N435

20	659.5	12.8	1229	5	Q94185	Q94185 caenorhabd1
21	657.5	12.8	2284	5	Q9VPG1	Q9VPG1 drosophila
22	653.5	12.7	1514	5	Q9GUM7	Q9GUM7 leishmania
23	647	12.6	4833	11	Q90YX6	Q90YX6 mus musculu
24	647	12.6	5038	11	Q90YX7	Q90YX7 mus musculu
25	641.5	12.4	2089	4	Q14676	Q14676 mus musculu
26	633	12.3	7489	10	Q41707	Q41707 homo sapien
27	632.5	12.3	7962	4	Q10465	Q10465 homo sapien
28	632	12.3	761	10	Q9ZQ10	Q9ZQ10 arabidopsi
29	629	12.3	6632	5	Q17362	Q17362 caenorhabd1
30	626.5	12.2	3507	5	Q23587	Q23587 caenorhabd1
31	626.5	12.2	990	13	Q91803	Q91803 xenopus lae
32	625	12.1	6642	5	Q01761	Q01761 caenorhabd1
33	624.5	12.1	1612	5	Q9VY02	Q9VY02 drosophila
34	624.5	12.1	2768	5	Q9VC00	Q9VC00 drosophila
35	622.5	12.1	839	2	Q9RX57	Q9RX57 delnoccoc
36	620.5	12.0	971	5	Q9XVS4	Q9XVS4 caenorhabd1
37	617	12.0	3570	4	Q99552	Q99552 homo sapien
38	609.5	11.8	801	5	Q9N3Y8	Q9N3Y8 caenorhabd1
39	607.5	11.8	2344	5	Q23635	Q23635 caenorhabd1
40	607.5	11.8	824	12	Q99307	Q99307 epstein-bar
41	605	11.7	379	5	Q27929	Q27929 drosophila
42	600.5	11.6	1893	5	Q9NKC9	Q9NKC9 drosophila
43	598.5	11.6	409	10	Q9SBM1	Q9SBM1 volvox cart
44	584.5	11.3	401	6	Q77765	Q77765 bos taurus
45	577	11.2	956	10	Q9LJ64	Q9LJ64 arabidopsi

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	1404 AA.
ID	Q92954			
AC	Q92954			
DT	01-FEB-1997 (TREMBLrel. 02, Created)			
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	MEGAKARYOCYTE STIMULATING FACTOR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
RA	Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,			
RA	Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
RL	Blood 78:279-279(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,			
RA	Fitzgerald M., Scalitro J., Kelleher K., Preissner K., Kriz R.,			
RA	Jacobs K., Turner K.,			
RL	(In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,			
RL	Mosher D.F. (eds.);			
RL	Biology of vitronectins and their receptors., pp.45-52,			
RL	Elsevier Science Publishers B.V. (1993).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
RA	Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,			
RA	Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.			
DR	EMBL, U70136; AB09089.1; -			
DR	InterPro: IPR000585; Hemopexin.			
DR	InterPro: IPR002402; Somatomedin_B.			
DR	InterPro: IPR002400; GF_cyskn.			
DR	Pfam: PF00045; hemopexin; 2.			
DR	Pfam: PF01033; Somatomedin_B; 2.			
DR	PRINTS, PR00438; GFCYSKNOT.			

DR PRINTS: PRO0022; SOMATOMEDINB.
DR PROSITE; PS00524; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Query Match	100.0%;	Score 5155;	DB 4;	Length 1404;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 968;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	VKDNKKNRKTKKKPRPPKPVYDEAGSGJDNDEKVFYTPDSTTQHNKVNSTSPKITTAKP	60
Db	200	VKDNRKNNKTKKPPKPPVYDEAGSGJDNDEKVFYTPDSTTQHNKVNSTSPKITTAKP	259
Qy	61	PRPSLPRNSDSKETSJLVNKEFTVETKEPTTNNKOTSFDEKERTSAKETOSIEKTSAK	120
Db	260	PRPSLPRNSDSKETSJLVNKEFTVETKEPTTNNKOTSIDGKRKKTISAETOSIEKTSAK	319
Qy	121	DLAPTSKVLAKPPEAKETTKGSPALTPKEPPTPEKREPASTTPKBEPTTTIKSAPTTK	180
Db	320	DLAPTSKVLAKPPEAKETTKGSPALTPKEPPTPEKREPASTTPKBEPTTTIKSAPTTK	379
Qy	181	EPAPTTTKSADPTPEKAPATTTKEAPATPEKAPATTTKEAPATTTKAPATTPKEPAPT	240
Db	380	EPAPTTTKSADPTPEKAPATTTKEAPATPEKAPATTTKEAPATTTKAPATTPKEPAPT	439

Query Match	99.8%	Score 5146;	DB 4;	Length 1404;
Best Local Similarity	99.8%	Pred. No. 0;		
Matches 966;	Conservative	0;	Mismatches 2;	Indels 0;
				Gaps 0.

Db	1100	AEGETPMLLRPHVEMEVETPPMDXYLPFRPNOSITINPALSDETINCNCKPVDGLITLRN	1155
Oy	961	GTIVAFRG 968	
Db	1160	GTIVAFRG 1167	
RESULT	2		
O9BX49			
ID	O9BX49	PRELIMINARY;	PRT; 1404 AA.
AC	O9BX49		
DT	01-JUN-2001 (TREMBlrel. 17, Created)		
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)		
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)		
DE	BGI74L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR).		
GN	BGI74L6.2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Wray P.;		
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; ALI3553; CAC36090.1; -;		
SQ	SEQUENCE 1404 AA; 151076 MW; 762A11746B3FDEE5 CRC64;		

Oy	301	PTTEPEAPTTKSPSTTKEAPATTKSGAPTTKEAPATTKSGAPTTKSPSTTKE	360
Db	500	PTTEPEAPTTKSPSTTKEAPATTKSGAPTTKEAPATTKSGAPTTKSPSTTKE	559
Oy	361	PAPTTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT	420
Db	560	PAPTTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT	619
Oy	421	TPKLTPTTPKLAPTTPEKAPATTPELAPTTPEEPTPTPEEAPATTPKAAANTKE	480
Db	620	TPKLTPTTPKLAPTTPEKAPATTPELAPTTPEEPTPTTPEEAPATTPKAAANTPE	679
Oy	481	PAPTTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT	540
Db	680	PAPTTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT	739
Oy	541	EPTSTSTKRAPATTPKGAPTTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT	600
Db	740	EPTSTSTKRAPATTPKGAPTTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT	799
Oy	601	KELAPTTTKGAPTTSDKAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT	660
Db	800	KELAPTTTKGAPTTSDKAPATTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT	859
Oy	661	PTTTEPTTIHKSPESTPELSAETPKALENSKEPEGVTTTTPAATKPEMTTKKDT	720
Db	860	PTTTEPTTIHKSPESTPELSAETPKALENSKEPEGVTTTTPAATKPEMTTKKDT	919
Oy	721	TEMDLTPTPETTTAPKMTKEATTTTEKTESKTTATTOVSTTODDTPFKITLAKT	780
Db	920	TEMDLTPTPETTTAPKMTKEATTTTEKTESKITATTOVSTTODDTPFKITLAKT	979
Oy	781	TLAPKVTYTTTKITTTTEIMNKEEETAKKADATNSKATTPPOKPTAPAKKPTSTKPKT	840
Db	980	TLAPKVTYTTTKITTTTEIMNKEEETAKKADATNSKATTPPOKPTAPAKKPTSTKPKT	103
Oy	841	MPRVKPKTPTPRKMTSTMPELNFTSIAIAMQTTRNQTPNSKLVEYNKSESDAG	900
Db	1040	MPRVKPKTPTPRKMTSTMPELNFTSIAIAMQTTRNQTPNSKLVEYNKSESDAG	109
Oy	901	ABSETPHMLRHVHPEVTPMDYLPRVNOGIIINPMLSDENINONKGPVNGIITLRN	960

Db	1100	AEGETPMLLRPHVEMEPVTPMDIYPRVNGIITINPLSDETNICNGKRVYDGLTTRN	1159
Qy	961	GTGVAFRG 968	
Db	1160	GTGVAFRG 1167	
RESULT	2		
ID	Q9BX49	PRELIMINARY; PRT: 1404 AA.	
AC	Q9BX49		
DT	01-JUN-2001 (Tremblrel, 17, created)		
DT	01-JUN-2001 (Tremblrel, 17, last sequence update)		
DT	01-JUN-2001 (Tremblrel, 17, last annotation update)		
DE	BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR)		
GN	BG174L6.2		
OS	Homo sapiens (human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OC	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Wray P.;		
RL	Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL; ALJ3353; CAC36090.1;		
SO	SEQUENCE 1404 AA: 151076 MW: 782A11746B3FDEE5 CRC04;		
Query Match	99.8%; Score 5146; DB 4; Length 1404;		
Best Local Similarity	99.8%; Pred. No. 0;		
Matches	966; Conservative 0; Mismatches 2; Indels 0; Gaps		
Qy	1	VKNKKNTKKKPPKPPVVDAGSGSDNGDFKVTTPDSTOHNKVSTSPKITTAKPIN 60	
Db	200	VKNKKNTKKKPPKPPVVDAGSGSDNGDFKVTTPDSTOHNKVSTSPKITTAKPIN 259	
Qy	61	PPSPSPSDSKETSILVNKEITVETKETTITNNKQTSIDGKEKITSAKETOSIEKTSAK 120	
Db	260	PPSPSPSDSKETSILVNKEITVETKETTITNNKQTSIDGKEKITSAKETOSIEKTSAK 319	
Qy	121	DLAPPSKVLAKPTKRAEITTTGAPLTTTPKEPTTPKEPASTPKKPEPTTIKSAPTPK 180	
Db	320	DLAPPSKVLAKPTKRAEITTTGAPLTTTPKEPTTPKEPASTPKKPEPTTIKSAPTPK 379	
Qy	181	EPAPTTKSAPTPKEPAPTTTKKEPAPTTKKEPAPTTTKSAPTTKKEPAPTTK 240	
Db	380	EPAPTTKSAPTPKEPAPTTTKKEPAPTTKKEPAPTTTKSAPTTKKEPAPTTK 439	
Qy	241	PKKPAPTTPKEPAPTTTKEPTTPTPKKEPAPTTKKEPAPTTKKEPAPPAKKPAPPTPKPA 300	
Db	440	PKKPAPTTPKEPAPTTTKEPTTPTPKKEPAPTTKKEPAPTTKKEPAPPAKKPAPPTPKPA 499	
Qy	301	PTTPKEPAPTTTKEPSPTPKKEPAPTTTKSAPTTTKKEPAPTTKSAPTTKKEPSPTTKE 360	
Db	500	PTTPKEPAPTTTKEPSPTPKKEPAPTTTKSAPTTTKKEPAPTTKSAPTTKKEPSPTTKE 559	
Qy	361	PAPPTPKKEPAPTTPKKPAATTPPKKEPAPTTTKKAPAPPAKKEPAPTTPKKEAPPT 420	
Db	560	PAPPTPKKEPAPTTPKKPAATTPPKKEPAPTTTKKAPAPPAKKEPAPTTPKKEAPPT 619	
Qy	421	TPKKLPTTPPEKLAATTPKEPAPTTPEELATTPPEELPTTPPEEPAPPTTKAAAPNTKE 480	
Db	620	TPKKLPTTPPEKLAATTPPEKAPATTPPEELATTPPEELPTTPPEEPAPPTTKAAAPNTKE 679	
Qy	481	PAPPTPKKEPAPTTPKKEPAPTTTKETAATTPPKGTAATTLKEPAPTTPKKPAKPELAPTTK 540	
Db	680	PAPPTPKKEPAPTTPKKPAATTPPKETAATTPPKGTAATTLKEPAPTTPKKPAKPELAPTTK 739	
Qy	541	EPSTTSDDPAPPTPKGTAATTPPKKEPAPTTTKKEPAPTTPKGTAATTLKEPAPTTKKKAP 600	
Db	740	EPSTTSDDPAPPTPKGTAATTPPKKEPAPTTTKKEPAPTTPKGTAATTLKEPAPTTKKKAP 799	


```

QY 601 KELAPTTTGGPTSTSDKAPPTPKETAATTEKPEAPPTPKKAPPTPEPTSEVT 660
DB 800 KELAPTTTGGPTSTSDKAPPTPKETAATTEKPEAPPTPKKAPPTPEPTSEVT 859
QY 661 PTTTKEPTTIHNSPDESTPELSAEPYKALENSPEKGVPTTKTBAATKPEMTTAKDKT 720
DB 860 PTTTKEPTTIHNSPDESTPELSAEPYKALENSPEKGVPTTKTBAATKPEMTTAKDKT 919
QY 721 TERDLRTPEPTTAAPKMTKETATTEKTESKITAATTOYSTTOOTPEKITTLLKT 780
DB 920 TERDLRTPEPTTAAPKMTKETATTEKTESKITAATTOYSTTOOTPEKITTLLKT 979
QY 781 TLAPRVTTTKITTTTEIMNKPDEETAKPKDRATNSKATTPKPKKPTSTKPKPT 840
DB 980 TLAPRVTTTKITTTTEIMNKPDEETAKPKDRATNSKATTPKPKKPTSTKPKPT 1039
QY 841 MPRVKKPPTTPPKMTSTMBELNPTSLIAEAMLOTTTRPQOTPNKSLIVEVNPSEADAG 900
DB 1040 MPRVKKPPTTPPKMTSTMBELNPTSLIAEAMLOTTTRPQOTPNKSLIVEVNPSEADAG 1099
QY 901 AEGETPHMLLRPHVFMPEVTPDMMDYLPRVNOGIINPMLSDETINICNGKRVODLTLLRN 960
DB 1100 AEGETPHMLLRPHVFMPEVTPDMMDYLPRVNOGIINPMLSDETINICNGKRVODLTLLRN 1159
QY 961 GTLVAERG 968
DB 1160 GTLVAERG 1167

RESULT 3
Q9JM99 PRELIMINARY; PRT; 1054 AA.
ID 09JM99:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE MNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN PRG4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikegawa S., Nakamura Y.:
RT "A novel mouse gene highly homologous to a human gene encoding
RT megakaryocyte stimulating factor precursor and cartilage superficial
RT zone protein."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB034730; BAA92310.1;
DR MGI; MGI:1891344; Prg4.
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR002365; P-rich_extensn.
DR Pfam; PF01033; Somatomedin_B; 2.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00202; SOMATOMEDINB.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;

```

Query Match 39.7%; Score 2046; DB 11; Length 1054;
 Best Local Similarity 45.9%; Pred. No. 3,le-123;
 Matches 449; Conservative 34; Mismatches 128; Indels 368; Gaps 20;

QY 1 VDNKKNRTKKRTPKPPVYVDEAGSLNDGDFKVT--TPDSTTOHKNVSTSPKTTAKP 58
 DB 195 VDNKKNRTKKRTPKPPVYVDEAGSLNDGDFKVT--TPDSTTOHKNVSTSPKTTAKP 254

```

QY 59 INRPSSLPPNSDSKESLTVNKEETVETKETTNNKOTSDOKERTTSKETS 118
DB 255 VTKPSLAPNSSETSKESLASNKEETVETKETTATNKOSA-SKKTTSVKETRSKETS 313
QY 119 AKDLAPTSKVLAPTPKAETTTGAPALTTPEKPTTPPKBPASTTKEEPTTKAGAPT 178
DB 314 KDOV-----EPSTYTK-----NSAPT 331
QY 179 PKEPAPTTKSAPTTKPEAPTTTKPEAPTTKPEAPTTKPEAPTTKSAPTTKPEAP 238
DB 332 TKRPV-TTTKESKFLP-----LQEPPTTAKKEPPTTKPEPTTKPEEP 376
QY 239 TTPKPKAPPTPKPEAPPTKEPTPTTPKPEAPTTKPEAPTTKPEAPTTKPKKAP 298
DB 377 TTPKPEEPTPKPEEPTTPKEEPTTPKPEEPTTPKPEEPTTPKPEEPTTPKPE 436
QY 299 PAPTTPKPEAPTTKPEEPTTPKPEAPTTKSAPTTKAPAPTTKSAPTTPKESPTT 358
DB 437 PEPTTKPEEPTTPKPEEPTTPKPEEPTTPKPEEPTTPKPEEPTTPKPEEPT 474
QY 359 KEBAPTTKPEAPTTPKKAPPTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 418
DB 475 KEBEPTTPKEEPTTPKPEEPTTPKPEEPTTPKPEEPTTPKPEEPTTPKPEEPT 512
QY 419 PTTPKLPTTPEKLAPTTPKPEAPTTPEELAPTTPEEPTTPPEEAPTTPKAAPMP 478
DB 513 -----TP 514
QY 479 KEBAPTTKPEAPTTPKPEAPTTPKETAATTPPKGTAATTKPEAPTTKKAAPKELA 538
DB 515 KEBEPTTPKEEPTTPKPEEPTTP-----KEEPTTPKPK----- 549
QY 539 TKEPTSTSDKAPPTPKGTAATTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPK 598
DB 550 -----EPTTPKEEPTTPKPEEPTTP-----KEEPTTPKEP 581
QY 599 AKELAPTTTGGPTSTSDKAPPTPKETAATTEKTESKITAATTOYSTTOOTPEKITT 658
DB 582 EP-----TTKKEEPTTPKPEEPTTPKPEEPTTPKPEEPTTP----- 617
QY 659 STPTTKEPTTIHNSPDESTPELSAEPYKALENSPEKGVPTTKTBAATKPEMTTAKD 718
DB 618 ----- 617
QY 719 KTERDLRTPEPTTAAPKMTKETATTEKTESKITAATTOYSTTOOTPEKITTLLK 778
DB 618 -----TSPKT-----TTLK 626
QY 779 TTTLAPKVTTTKITTTTEIMNKPDEETAKPKDRATNSKAT-----TKRP-OKPTKAP 829
DB 627 ATTLAPKVTAPE-----EIONKKEEPTTPASESDSDSKTTLKPKPKAKPKPKPKAP 681
QY 830 KKPSTSKPKKMPRVKPKKPTTPPKMTSTMBELNPTSLIAEAMLOTTTRPQOTPNKSLV 889
DB 682 KKPSTSKPKKPT-PKTRKPKTPPALKTTSATPELNTTP--LEVLWPTPTTTPKQTPNETA 738
QY 890 EVNPKSEADAGAEGETPHMLLRPHVFMPEVTPDMMDYLPRVNOGIINPMLSDETINICNG 949
DB 739 EVNPDHDDADOGEBEKP-LRGPVLFPTALPGDLAGLRNGININPMSDETINICNG 797
QY 950 KPVVGLTTLRNGTLVAERG 968
DB 798 KPVVGLTTLRNGTLVAERG 816

```

RESULT 4
 Q9N4S7 PRELIMINARY; PRT; 1079 AA.
 ID Q9N4S7:
 AC Q9N4S7:
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE Y51B1A.1 PROTEIN.

GN *Y51B1A.1*.
 OS *Caenorhabditis elegans*.
 OC *Pukayoya; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;*
 OC *Rhabditidae; Peloderinae; Caenorhabditis*.
 NX NCBI_TaxId=6239;
 RX [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;
 RC MEDLINE=99069613; PubMed=9851916;
 RX None;
 RA "genome sequence of the nematode *C. elegans*: a platform for
 RT investigating biology. The *C. elegans* Sequencing Consortium.";
 RT Science 282:2012-2018(1998).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;
 RC Cotton M.;
 RA "the sequence of *C. elegans* cosmid Y51B1A.";
 RT submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
 RL [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC006797; AAF60743.1; "
 DR InterPro: IPR002965; P-rich extensin.
 DR PRINTS: PR01217; PRICHEXNSN.
 SO SEQUENCE 1079 AA; 110532 MW; 8BDBE3824CF80CA1 CRC64;

Query Match	18.8%;	Score 967;	DB 5;	Length 1079;
Best Local Similarity	32.3%;	Pred. No. 2.9e-54;		
Matches 294;	Conservative 114;	Mismatches 377;	Indels 124;	Gaps 33

[illegible][illegible]

RESULT	5	
091751	PRELIMINARY;	PRT: 1049 AA.
ID	091751	
AC	091751	
DT	01-MAR-2001 (TREMblrel. 16, Created)	
DT	01-MAR-2001 (TREMblrel. 16, Last sequence update)	
DT	01-JUN-2001 (TREMblrel. 17, Last annotation update)	
DE	CG5228. PROTEIN.	
GN	CG5228.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BERKELEY;	
RX	MEDLINE=20196006; PubMed=107311132;	
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yeandle M.D., Zhang O., Chen L.X.,	
RA	Bratton R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,	
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Aburat F.F., Aburat A., An H.-J., Andrews-Pfankoch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies S.M.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Foster C., Gabrielian A.E., Garg N.S., Guan P., Harris M.,	
RA	Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,	
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kemsion J.A., Kethum K.A.,	
RA	Kimmel B.E., Kodera C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,	
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., Moberg D.,	
RA	Merklow G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,	
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,	
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,	
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	
RA	Reinelt K., Remington K., Saunders R.D.C., Scheele F., Shen H.,	
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skipski M.P., Smith T.,	

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang R.A., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003495; AAC22353.1;
 DR Flybase: FBgn0030561; CG5228.
 SO SEQUENCE 1049 AA; 107278 MW; 954DD629E7619671 CRC64;

Query Match 18.2%; Score 936.5; DB 5; Length 1049;
 Best Local Similarity 34.6%; Pred. No. 2,5e-52;
 Matches 371; Conservative 45; Mismatches 439; Indels 217; Gaps 54;

QY 2 KONKNNRTKKKPPKPPVVDGSGLDNGDKRVTPDTS---TQHNKVSPTSRTTAKP 58
 DB 30 KNEIGTAKPTTLKP-----TEGSAKPTTLKPTEGSAKPTTLKP 71
 QY 59 -----INRPSLPNDSITSKESL-----TVNKETVTEKETTNNQOT---STGKEVTT 106
 DB 72 TEGTAKPTTLKPTEGTAATTLKPTEGSAKPTTLKPTEGTAATTLKPTEG---T 127
 QY 107 SAKET--OSIEKTSK--DLAPTSKVLAKPT-----PKAETTKGPAITPKPEPTPT 155
 DB 128 SAKPTTLKPTEGTAATTLKPTEGSAKPTTLKPTEGSAKPTTLKPTEGSAKPTTLK 187
 QY 156 PKE-----PASTTPKEPT---PTTKSAKPTTLKPTEGSAKPTTLKPTEGSAKPTTLK 207
 DB 188 PTEGSAKPTTLKPTEGSAKPTTLK--PTEGTAATTLK--PTEGSAKPTTLKPTEG 243
 QY 208 TTPKEPAATTKKEPAATTKSAAPTTPKE--PAPTPKKPAATTKKEPAATTKPEPT---P 262
 DB 244 TTAK--PTTLKPTEGTAATTLKPTINPEGSAKPTTLKPTEGTAATTKPTINPEGSAK 300
 QY 263 TTPKEPAATTKKEPAATTKKEPAATTKKPAATTKKEPAATTK-----EPAPTTKPEPT 317
 DB 301 TTKLPTEGTAATTLKPTEGSAKPTTLKPTEGTAATTKPTTLKPTEGSAKPTTLKPTD 360
 QY 318 TTPKEPAATTK-----TKSAATTKKEPAATTKSA--PTTPKEPSPTTKKEPAATTKPE 368
 DB 361 TTAK--PTTLKPTEGSAKPTTLKPTEGTAATTKPTTLKPTEGSAKPTTLKPTEGTAAT 416
 QY 369 PAPTPKKPAATTKKEPAATTKKEPAATTKKPA--PTAPKEPAATTKKEPAATTK--PK 423
 DB 417 --PTTLKPTEGTAATTK--PTTLK--PTEGTAATTKPTTLKPTEGSAKPTTLKPTEGTAAT 470
 QY 424 KLPPT-----TPEKLAATPEKAPATTPBELAPT--TPEEPTPTTPEEPAATTKKAAAPN 476
 DB 471 TTKPTEGTAATTKPTTLKKTGKTSKAPTTLPTEGTAATTKPTTLKPTEGTAATTKLP 530
 QY 477 TTPKEPAATTKKEPAATTKKEPAATTKETAPTPKSTA-----PTTLKEPAATTP 526
 DB 531 KGTAKTAATTLK--PTEGSAKPTTLKPTEGTAATTKPTTLKPTEGSAKPTTLKPTEGTAAT 587
 QY 527 K-----KPA-----PKELAPT-----TKKEPTSTTSKPA-----PTTPKGT 558
 DB 588 KPTTLKPTEGSAKPTTLKPTEGSAKPTTLKPTEGTAATTKPTTLKPTEGSAKPTTLKPT 647
 QY 559 APPTPEPAATTKKEPAATTKPKGTAPTTLPKEPAATTKPK-----KPAKELAPTTTKGPTST 614
 DB 648 EGTTAK--PTTLK--PTEGSAKPTTLKPTEGTAATTKPTTLKPTEGTAATTKPTTLKPTEG 701
 QY 615 TSKKPAATTKPK--TAPTPKEPAATTKPKPA--PTPEPTPTTSEVSTPT--TTKEPT 668
 DB 702 TSKAPTTLPTEGTAATTKPTTLKPTEGSAKPTTLKPTEGTAATTKPTTLKPTEGTAATTK 761
 QY 669 TTKSDESTPELSAETP-----KALENSPKPEGV--PTTK--PAAATK 709
 DB 762 TL--KPEGT--SAKPTTLKPTEGTAATTKPTTLKPTEGSAKPTTLKPTEGSAKPTTLK 816

QY 710 PEMTTAKDKTTERDLRT-----PENTTAAPKMKETATP-----TEKTTES 752
 DB 817 PTEGTAATTKPTTLKPTEGSAKPTTLKPTEGTAATTKPTTLKPTEGSAKPTTLKPTEGTAAT 876
 QY 753 KITAATTTQVSTTQDTPPRKITTLLKTTTAPKVTYTKK--TTTTEIMNKPEETAKPKD 810
 DB 877 PTTKLPTEGSAKPTTLKPTEGTAATTKPTTLKPTEGTAATTKPTTLKPTEGSAKPTTLKPT 936
 QY 811 RATNSKATTPKPOKPTKAPK--KPT--STKKPKTPRVKPKPTTP--PRKMTSTMP- 861
 DB 937 -GTAKPTTLKPTEGSAKPTTLKPTEGTAATTKPTTLKPTEGSAKPTTLKPTEGSAKPTTL 995
 QY 862 ELNPTSIAEAMQTTTPNQTNSKLVENPKSEDAAGC-ETPHMLRP 912
 DB 996 TTKPTEGSAKPTTLKPTEGTAATTKPTTLKPTEGSAKPTTLKPTEGSAKPTTLKPT 1045

RESULT 6
 076894 PRELIMINARY; PRT; 1795 AA.
 AC 076894;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE EG:56G7.1 PROTEIN
 GN EG:56G7.1 OR CG14796.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 NC NCB1_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
 RA Shue B.-C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang R.A., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).
 RN
 RA SEQUENCE FROM N.A.
 RA Caden E., Dreano S., Lelaure V., Mottier S., Gilbert F.:
 RT "Sequencing the distal X chromosome of Drosophila melanogaster."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Benos P.:
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003421; AAF4564.1; -
 DR EMBL: AL031028; CAI19845.2; -
 DR FlyBase: FBgn0025390; EG:5667.1.
 DR InterPro: IPR002557; Chitin_binding.
 DR InterPro: IPR002965; P_rich_extensn.
 DR Pfam: PF01607; Chitin_bind_2; 2.
 DR PRINTS: PRO1217; PRICHEXTENS.
 DR SMART: SM00494; CNEBD2; 2.
 SO SEQUENCE 1795 AA; 194464 MW; 07F10C129BD9557B CRC64;

Query Match 16.8%; Score 865.5; DB 5; Length 1795;
 Best Local Similarity 28.5%; Pred. No. 1.4e-47;
 Matches 276; Conservative 129; Mismatches 447; Indels 115; Gaps 28;

QY 6 KNRTKKRPKPPVVDAGSGLDNGDKVTPDTSTTHNNKSTSPKITTAKPLNPKPSL 65
 Db 431 KNDYTEAPEIKSP-----KGLHSENIIVILPETTTT--TTTTTKPVVLCPTISPDDT 483
 QY 66 PPNSDT-----SKESLVNKEETVETKETTNNKSTGDKKETSASKEQSI 114
 Db 484 PKPSTTTAVTKSPKISSTEOHSTTKAKTTTKKPTVTEKTSATKRPRTVTTTOK 543
 QY 115 EKTAKDLAPSKVLAKTPKAEETTTGKALTTKEPEPTTPKPEASTTKEPTTIKS 174
 Db 544 RSTTHNSPTKTTINSTLSPKTTTTPSTTTPSTTSPST-----TTPSTTTPST--- 594
 QY 175 APPTPKPEAPTTTASAPTTKEPAATTK--EPATTPKEPAATTTKBPATTTKSAPTTP 233
 Db 595 --TTPSTTTPSTTTTAVKSTHRPRTSOKTTTASTTTK---TTTSBKTTKTIDTPTST 648
 QY 234 KEPAATTKKRPAPTPKPEAPTPKEPTTPKBPATTPKBPATTPKBPATPKKAP 293
 Db 649 TSKLSTTOKTTTTHKTAATSTERKKTTEKTSVSTTKKSTESSPKPTSTGKPT 708
 QY 294 TTPKPEAPTPKPEAPTTTKEPSPTTPKPEAPTTTKSAPTTPKBPATTTKSAPTTPKEP 353
 Db 709 TTPKSTRTTPTTTPKVTITTOITTTTTPLRSSNETTSTOPEPTTTTTPQPTTTTLVTPKTS 768
 QY 354 SPTTKEPAATTPKPEAPTPTPK-----KBPATTPKEPA-----TTPKE 392
 Db 769 TTTTTEKAPITSSPKPTTTTOKITSTANTKVAITTOKEPTTPQOSTTITFTKTTTNN 828
 QY 393 PAPTTTKKRPAPTPKPEAPTPK--ETAPTPPKKLTTPPEKIAPTTPKAP-----AP 443
 Db 829 TTPSTTEKRTTTPKPTSTPKSTIVASSTETKTISSPKPTTEKSTNPTNSVKTAL 888
 QY 444 TTPPELAPTPPEPTTPPEEPAPTPPRAAANPKBPATTPKBPAPT--TPKBPAPT 501
 Db 889 TSTGATSTSEPTKKT--ONITTTTPKPTTIKISTOEATITSTQKSVITTTKATSSS 947
 QY 502 PKETAPTPPKGAPTTTKEPAATTPKPKAPKELAPTTTKEPT--STTSDBKAPPTPKGTA 559
 Db 948 PLTILSTEEPNTPKPLRTTPTTTSVATRTITTTTISESTETTSOKPKSTPTSTT 1007
 QY 560 PTTTKEPAATTPKPEAPTPPKGAPTTTKEPAATTPKPKAPKELAP--TTTKGPTSTTSOK 618
 Db 1008 RTTPKVTIVIVSTQNTPTTTSKISTVTI-----TTP--NESPSTQRTTTPKPTISITAST 1061
 QY 619 PAPTTTKKAPTPKPEAPTPPKKAPTPPETTPPTTSVSTPTT--TKE-----666
 Db 1062 TSGTTRITPTTNPONSTSTDLITVTRKPCDPBSTSKNTNINACTOQLQOVNILELOS 1121

QY 667 -----PTTIHKSPPDESTPELSAEPTRKALENSPEKPGVPTTPATKPEWTTAKD 718
 Db 1122 POKOQFHTHTHTALGNSNTLGGQEVDPYMDAPSSAEPESQATTAAPMTSTLAA 1181
 QY 729 KTER---DLRTPEETTAAPKMTKETATTTTEKTESKITATTTQVSTTODTTPKIT 775
 Db 1182 HLOKLFHISTTPSRHAP--TORPSSQSSSORSR--GVTIQMARHNLIASKPTAH 1238
 QY 776 TTKTT---TLAPKVTYTKKITTTTTEINMKPEETAKPKRDATNKAATTPKOKPTKA--PK 830
 Db 1239 SLRLSIQOLASTOKRSIPKTLVYHNTKPEP--SEYIDSETSIQYIDEDNEVLDTQPR 1237
 QY 831 KTSSTKKRTPMPRVKPKPTTPPKRMTSTMPELNPTSRIAEAMLOTTTPRNOTPNSKIVE 890
 Db 1298 AMSSTTVAALVAPVSTTTERPK--TSSP--SPT-----KATSETTQPIETTTGDLLEY 1350
 QY 891 VNPKSED 897
 Db 1351 DSSGSSD 1357

RESULT 7
 ID 096449 PRELIMINARY; PRT; 1489 AA.
 AC 096449;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
 DE CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
 GN CAR90.
 OS Phytophthora infestans (Potato late blight fungus).
 OC Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
 OC Phytophthora.
 NCBI_TaxID=4787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RACE 1-11;
 RA Goerhardt B.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF061185; AAC72308.1; -
 SO SEQUENCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;

Query Match 16.5%; Score 853; DB 10; Length 1489;
 Best Local Similarity 32.3%; Pred. No. 7.3e-47;
 Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;

QY 9 TKKKRPKPPVVDAGSGLDNGDKVTPDTSTT-----OHNNKSTSPKITTAKPINR 62
 Db 338 TTPAFTKRPYDVEETTYVTEESTYAPTCKSETNAPTEBMVYAHIEKPCDTEVMYAPTEET 397
 QY 63 PSLP-----PNSDT-----SKESLVNKEETVETKETT---TNKQSTND 100
 Db 398 TYAPTEETTYAPTEETTYAPTEETPYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 457
 QY 101 GKEKTTAKETQSEKSAADLAPTSKVLAKTPKAEETTTGKALTTTPKEPTTPKEPA 160
 Db 458 TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 514
 QY 161 STTPKEPTPTTIKSAPTTPKPEAPT--TTKSAPTTPKPEAPT--TTKEPAATTPKPEAPT 217
 Db 515 EETTYEPTTEET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 573
 QY 218 TKEPAATTTTSAPT-----TTPKPEAPTPPKKAPTPKPEAPTPPKPEPTTPTPKE 267
 Db 574 TYAPTEETTY--YAPTEETMYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 632
 QY 268 P-----TTKEPAATTPKPEAPT-----APAKKAPATTPKPEAPTP 304
 Db 633 PTEETTYASTEETTYAPTEETTYAPABETPEPEETTYAPTEETTYAPTEETTYAPTEET 692
 QY 305 KEPAAT--TTKEPSPTTPKEP-----APT--TTKSAPT-----TKEP-----A 339

Db 693 TTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETMYAPTEETTYGPEETTYA 752
 QY 340 PT-TTKSPAPT--TPKPSPTTKEP-----APTTPKEAPPTPKKAPPTTKKEAP-- 388
 Db 753 PTEATTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEE 812
 QY 389 TPKEAPPTTKKAPAPAPKEAPPTTKETAPTPPKLAPPTPEKLAAPTPEKAPTEEE 448
 Db 813 TPEEPTTEETTYAPTEETPYEPTTEET--TYTPTEETTYAPTEETTYAPTEETTYAPTEET 870
 QY 449 LAPT--TPEEPTPTPEEP-----APT-----TPKAAANTPKKEAPPTPKEP- 489
 Db 871 YAPTEETPYEPTTEETTYAPTKETTYAPTEETTYASTEETTYAPTEETTYAPAEPTPEPT 930
 QY 490 -----APTTPKEAPPTTKETAPTPPKGTAP--TLKEAPPTPKAPK-ELAPT--TT 539
 Db 931 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEET 990
 QY 540 KEPTSTSDKAPAPTPKGTAPPTPKKEAPPTPKEP-----APTTPKGTAPT----- 585
 Db 991 YAPTEETMYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAST 1050
 QY 586 -----TLKEAPPTPKKP-----APKE--LAPT--TTKGPSTSDKAPAPTP 624
 Db 1051 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1110
 QY 625 KETAPTKKEAPPTPKKAPAPTPPEPTTSEVSTPTTKETP--TIKSPDESTEPLS 682
 Db 1111 YAPAEETTYEPTTEETTYAPTEETTYAPTEETMYAPTEETTYGPEETTY-APTATTYAP 1169
 QY 683 AEPTPKA-LENSPKEPG-----VPTTKPATKPEMTTKAKDKTEEDLRTTP---ETT 732
 Db 1170 TEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1229
 QY 733 TAARKMKETVATTEKTESKITATTOVSTTQDTT-----PEKITP---LKT 779
 Db 1230 TYAP-----TEETTYEPTTEETTYAPTEETTYAPTEETTYAPTEETMYAPTEETTYGTEE 1284
 QY 780 TPLAPKVTTKITTTTEIMNKKPEETAKPKDRAT--NSKATTPKP-OKPTKAPKKPT-- 833
 Db 1285 TTYAPTEETTYAPTEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPAEETPYE 1344
 QY 834 -----STKRP-KTMPRVARKPTTPTRKMTSTWELNPTSRLEAALQTTTP--N 881
 Db 1345 PAESTSTVSTKEKPCNTEETDEPTDEPTDE--PSDEPTDEPTDEPTDEPTDEPTPCDN 1402
 QY 882 QTPRSKLVEVNPKESEAG 899
 Db 1403 QGINGIGVENKVRYNAG 1420

RESULT 8
 Q9SPMO PRELIMINARY: PRT: 1315 AA.
 AC Q9SPMO;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE EXTENSIN-LIKE PROTEIN.
 GN PEX2.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=POLLEN;
 RA Stratford S., Barnes W., Golubiewski A., Colter R., McCormick S.,
 RA Honorst D., Gao M., Shewalter A., Bedinger P.A.;
 RT "Polen Extensin-like (Pex) Genes in a Monocot and a Dicot."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF159297; AAD5980.1; -

DR InterPro: IPR001611; LRR.
 DR InterPro: IPR00198; Xylose_isom.
 DR InterPro: IPR002965; P-rich_extensin.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 3
 DR PRINTS: PR01217; PRICHEXTENS.
 DR PROSITE, PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
 DR SMART: SM00370; LRR; 3.
 SQ SEQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;

Query Match 15.7%; Score 808.5; DB 10; Length 1315;
 Best Local Similarity 25.0%; Pred. No. 4,6e-44;
 Matches 255; Conservative 88; Mismatches 459; Indels 219; Gaps 33;

QY 13 PTP-----KPPVDEAGSGLDNGDEKVTPTPTSTTHNNKYSTSKITAPINR 62
 Db 412 PTPGPGSSSVPGKPPSVG-----KPAAPAPMTPTPTPDVSP-----EPL-PE 455
 QY 63 PSLRP-----NSDTSKETSLTVNKETTYVERKETTITNKOTSDGKEKTSAKETQ 112
 Db 456 PSEVPAPAPKMPMLKSPADEYIP--PVPAPKSPGTSPPASRCAPLQAOAPPAAS 511
 QY 113 SIEKTSKDLAPTSKY-----LAKPTPKATITTKGRALITPKKEPTTPKE 158
 Db 512 SPAPATVKSPPPAVAVLPPPAKTPSPAPVSPPEAPVSPQOVKSSPPPAVASP 571
 QY 159 PASTTKEPTPTTKISAPPTKEAPATTKSAPTP--KEAP-----TTTKEAPPTP 210
 Db 572 PPKSP--PPPAVAVSPPMLKSPPPAPVASPQPLKSPPPVLMSTPSKSSPPVP 629
 QY 211 KEAPATTKKEAPATTKSAPTPKE-----PAP--TTP-----KPAPTPKKEAPPTPK 259
 Db 630 VASPPPVKSPPLAPVSSPSPVKLPLPAPKSTPPEEKPPPPVKSPPPEKSL 689
 QY 260 PPTTPKEAPATTKKEAPATTKKEAPATKAPKAPPTPKKEAPATTKKEAPATTKESP 319
 Db 690 PPTLTTPSPPEKPTPTSTESKPPPPSPVETLPPSKSSPEEVSPPPAVSSSP-- 747
 QY 320 PKKEAPATTKSAPPTTKKEAPATTKSAPTPKEPSPTTKKEAPATTKKEAPATTKKAP 379
 Db 748 ---PAPVS--SPPLKSSPPPVSESSPPPTKSSPLAPVSSPPQVEKTSPPPAVSSP 802
 QY 380 TTPKEAPATTKKEAPATTKKAPAPKKEAPATTKKEAPATTKKLPPTPEKLAAPTPPE 439
 Db 803 PTPKSSPLAPVSSPPQVEKTSPPPAVSSPPPKSSSPLAPVSSPPQVEKTSPPPAV 862
 QY 440 KPAPTPEELAPTPPEEPTPTTPEEAPPTKKAAPNTPKKEAPATTKKEAPATTKKEAP 499
 Db 863 SSPLEPKRSSPPSSVSPPTTVKSSPPPALSSPPMTPKSSPPPAHVSSPPEAKSSPP 922
 QY 500 TTPKEATTPKGT-----PTTKEAPATTKKAPKELAPTTTKEPTSTSD 548
 Db 923 LAPSSPSEKSPSSPPVMEKTSPPPAIVSSPPPTKSSPP--APVSSPPVVKSP 979
 QY 549 KPAPT-----TPKGTAPTP-----KEAPATTKKEAPATTK--GTAPTLKEAP 593
 Db 980 PPAPVSSPPPTKPLPPAPVSSPPVVKSSPPPTTVSSPPPTPKLPPPTTVSSPPPT 1039
 QY 594 TPKKAPKELAPTTTGTSTSDKAPPT--PKETAPTPK--EPAPTTKKEAPPT 647
 Db 1040 KKPLPP--APVSSPPVVKSSPPPAVSLPPTTRKSPPTRVSSPRAVVCCEPPTL 1096
 QY 648 PETPPTTSEVSTPTTKETPTTIKSPDESTEPLSAPPEKALENSKEGCVTTTTPAA 707
 Db 1097 VSSPPAPKSLPPTTVSSP-----PEVKSSPPPTTVSSPPAPKSSPPPTPV 1146
 QY 708 TPKEMTTAKDKTTERDLKTP-----ETTTAAPKMTKETATTEKTESKITATTOYT 762
 Db 1147 SPP-----ELKSSPPPAVSSPPASAPKSSPPPAVSLPPEVKSSPPPAVPS 1194
 QY 763 STTQDTTPFKITLTKTTTILAPKVTTTKITTTTEIMNKEETAKKDRATNSKATTPKP 822

```

Db 1195 SPPPPAKSP-----PPPA-----PMSLPPVKSPPPPAVSSP 1228
QY 823 OKETAPKPKPTSTKPKTMPRVKPKETPTTPPKKATSTMPLELPTSRIAEMLOTTTRPNO 882
Db 1229 PPKMSKPPPPAPSSPPPPA--VKPPSLPPPPAVSSPPPAV-----TSAP-- 1271
QY 883 TPKSKLVEVNPKESEDGAGETPHMLLRPHVMPVETPDMDYLPFPVNOGIIINPMSD 942
Db 1272 -----PKKEEDSTA-----PPAELPPPPSNDIILPPIMAN 1302
QY 943 E 943
Db 1303 K 1303

RESULT 9
Q20007 PRELIMINARY; PRT; 1274 AA.
ID Q20007;
AC Q20007;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE COSMID F35A5.
GN F35A5.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Cretton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Lattelle P.,
RA Lightung J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showman R.,
RA Smailson J., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RA *2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RI Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Leimbach D.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U46675; AAB52641.1;
SQ SEQUENCE 1274 AA; 138065 MW; 07F6BD292C1799F CRC64;

Query Match 15.6%; Score 802; DB 5; Length 1274;
Best Local Similarity 28.0%; Pred. No. 1.2e-43;
Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;
QY 6 KNTTKK-KP-----TP-----KPPVVDGAGSLDNGD-----FKVTTPTDSTTQHNVSTS 50
Db 274 KNTTKKRPWEDETVVEEVEKPPVPEKAPVLKKKDPAPAKARDPSPSKAAPKKEVPS 333
QY 51 PKITTAKP1-NPPPSLPPNSDTSKETSIVNKKETTVETKETTITTTNKQSTSDGKEKTTSAK 109

```

```

Db 334 SPVVPPTPVKNPKYKIPPEVEDEPAEVEVKKSAPPEKTPVLKRNKEPSSSTPPSSDS 393
QY 110 ETOSIEKTSKADLAPTSKVLAKPTKAE-----TTTGA-----LTTP 148
Db 394 PKKAPVAKPRDSSPKKATPLQADPAKQEVPPPVKVPVKKRPWEVDEDEPVEEVOQ 453
QY 149 KEPITTP-----KRPASTPKEP-----TP--TTIKSAPTTKEPAPTTTSKAPTTKEPAP 199
Db 454 EAPAKKTPVLKRPKAPKDTAKPATSTPTPEKPKDDVKKRDSSPKVAAPKPSQAAPA- 512
QY 200 TTKKPA-----PTTKKAPTTTKKAPPT-----TKSAPTT 233
Db 513 TPVKNPVKMKRPWEDETPADVSKPTDAKKTPTSLAKKDPAPAKESLAKRADTKAPAKP 572
QY 234 KEP-----APTPPKKAPATTPKE-----PAPTPKEPTTPPKKAPTTKEPAPTTK-EP 283
Db 573 RDPSPKKVAPAPAKETTVLAKKBPAGADSKTEPEKSKPRDSSPKKAPVAPKTEV 632
QY 284 APAPKAP-----TPPKKAPTTKEPAPTTTKEP-----SPTTKKAPTTKAPPT 333
Db 633 APAAVKKKEPISKPDTPAKKAEKNSPVV--PTVKNPVKMKRPWEDEDDAPAKVSLPE 691
QY 334 TTKKAPTTTSKAPTP--KEPSPTTKKAPATTPK--EPATTPKKAP--TPKKEPAPT 388
Db 692 PEKK--TPVLAKKAPYKPDSEAAADPVSGPSSKDEPLAKKAVKPRDSSPKKAVDIFKAPK 750
QY 389 TTKKAPTTTKKAPTA--PKKAPTTTKETAATTPPKKLTPTTP-----EKLAPTTKPA 442
Db 751 T--EVPVAVKKEPVAKSRDSSPKKAK--AEKNS--VVPPTPVKNPVKMKRPWEDDA 805
QY 443 PTPPELAPTTKEEP-----PTTKEPAPTTPVKAAANTKEKAPTTKAPATTP--PKE 496
Db 806 PAEPVAVPEPEKTPVLAKKTTPVKKPRDSSPKKAVPAKSIKIDAPVAVKKEPESKKE 865
QY 497 PAPTTKEKAPTTPKGTATPTTKKAPATTPKPKAPKELAPT--TTKEPTSTSDKAPATTP 555
Db 866 PSKKAKEPSVVP-----PTPVKNPVKMK--KRPWEDEDEPTTEVKKSE--PEKTPVLA 918
QY 556 KGTAPPTKEPAPTTKREAPTTPKGTATPTTKKAPATTP-----TPKKAP--KEL 603
Db 919 K-KEPKKD-APKVAKPRDSSPKKAVE--KEPAKVAKPRDSSPKKALPDPANTQEA 974
QY 604 APPTTKGPS-----TSDKP-----APTTPKEP-----APTTPKEPAPTTKAPAT 646
Db 975 PPTPVKNPVKMKRPWEDEDEPAEVSAPPEKKTPTVLAKKAPAPROD--SPKKAAPV 1031
QY 647 TPTPTPTTSEVSTPTTKETPTIKSP-----DESTPELSA-EP--TPKALENSPKRPG 698
Db 1032 AAK-PDCKIPEV--PTPVKNPVKMKRPWEDEDEPVSAPPEKKTPTVLAKKAPTPA 1089
QY 699 V-----PTTKTPAAI-----KREMTAKKDTTERDLRTTPETT--TAAPK 737
Db 1090 TKPDESAADPVSGPTSKDLSKKAPVEKPTTPPKDOKLPSAPKAPKAPAPKAPK 1149
QY 738 MTKETATTTKTESKLTATTOVTSTTQDTPTPKITTLKTTLAPKVTTKTITTE 797
Db 1150 KWKPVWDDDEPADTVPAPSKKPTDEDAPPLG-----GPKTKDKP----- 1193
QY 798 IMNKPEETAKPDATNSKATTPKPOKVPYAKPKPSTKPKTMPRVKPK----- 848
Db 1194 -LNKKAAPAEKPTK-----PKPEVSKPEKPTPEPPK--AAPKMKRPWEDEDEPE 1243
QY 849 ---TTTPPKKTTSTMP 861
Db 1244 ADFTMPAKKPTDTP 1259

RESULT 10
Q9LIE8 PRELIMINARY; PRT; 1480 AA.
ID Q9LIE8;
AC Q9LIE8;
DT 01-OCT-2000 (TEMBLrel. 15, Created)

```


QY 129 -----LAKPTKAEETTTGPALETTKEPTTKEPASTTKEPTTKEPTTKEPA 183
 Db 542 SPPPPKSPPPAPVGSPPPPKSPPPAPVASSPPPPKSP--PPPTLVASPPPPKSP 599
 QY 184 PTTTKSAPTTT--KEBAPTTEKAPATTKEBAPTTEKAPATTTKSAPTTT--KEBAPT 240
 Db 600 PAPAVASPPPPKSPPPPPVASSPPPPVASSPPPPKSPPPPPVSSPPPPKSPPPPP 659
 QY 241 PKKPAATTTKEBAPT-----PKE---PTPTTKEBAPTTEKAPATTKEBAPTAPK 291
 Db 660 PAKSTPPPEEYTPPTSVKSSPPPEKSLPPPLISPPPOKEPTTPTSPK--PSSPEKP 718
 QY 292 APTTKEBAPTTEKAPATTKEPTTKEPTTKEBAPTTEKAPATTTKSAPTTTKEBAPT 351
 Db 719 SP--PKEPVSSPPQTP-----KSSPPAPVSSPPPTVSSPPALAVSSPPVSSKSP--- 768
 QY 352 EDSPTTKEBAPTTEKAPATTKEBAPTTEKAPATTKEBAPTTEKAPATTKEBAPT 411
 Db 769 PPAPLSSPPAPVOKSSPPPVVSSP--PPAPKSSPPALAVSSPPPOVEKTSPPALSSPP 827
 QY 412 TTPKETAP-----TTP-----KLLPTTPEKLAAPTTEKAPAP-----TTPBELA 458
 Db 828 LAPKSSPPHVVSSPPPVVKKSSPPAPVSSPPLTEKAPAPVSSPPVVKSTP--PA 885
 QY 459 PTTPEEAPTTTAKAANTPKE--PAPTTEKAPATTKEBAPTTEKAPATTKEBAPT 517
 Db 886 PTVIVISP--PSEPKSSPPPPVSLPPIVKSPPAPVSSP--PMTPKSSPPVAVSSPPT 943
 QY 518 LKE---PA-----PTTPPK--PAPKELAPTTTKE--PTSTSDAPATTKEBAPT 563
 Db 944 VKSSPPAPVSSPPATPKSSPPAPVNLPPPEVKSSPPTPVSSPPPA---PKSSPPAP 1000
 QY 564 --KEBAPTTEKAPATTKEBAPTTEKAPATTKEBAPTTEKAPATTKEBAPT 622
 Db 1001 MSSPPPEVKSPPPPAVSSPPPPVKSPPPPAVSSP--PVPKSSPPAPVSSPPPV 1057
 QY 623 --TKETAPTTKEBAPTTEKAPATTKEBAPTTEKAPATTKEBAPTTEKAPATT 681
 Db 1058 KSPPPAPVSSPPPPVKSPPPPAVS--SPPPPKSSPPAPVSSPPPKSPPAPV 1115
 QY 682 SAETTPALENSPKEGVTTKTPAATKPMATTAKDKTTERDLRTTPT 731
 Db 1116 SPPAP-----VKPPLP--PPAVSSPPPVVTPAPPKKEQSLPPAES 1158
 RESULT 12
 P70670 PRELIMINARY; PRT; 2187 AA.
 ID P70670: AC P70670:
 DT 01-FEB-1997 (Triblrel. 02, Created)
 DT 01-FEB-1997 (Triblrel. 02, Last sequence update)
 DT 01-JUN-2001 (Triblrel. 17, Last annotation update)
 DE NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
 MUSCLE-SPECIFIC FORM GP220).
 CN NACA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96312450; PubMed=8698236;
 RA Yotcov M.V., St-Arnaud R.;
 RT "Differential splicing in of a proline-rich exon converts alphaNAC
 RT into a muscle-specific transcription factor.";
 RL Genes Dev. 10:1763-1772(1996).
 DR EMBL: U48364; AAB18734.1; -;
 DR EMBL: U48363; AAB18732.1; -;
 DR MGP; MGI:106095; Naca.
 DR InterPro; IPR002715; NAC.
 DR InterPro; IPR003037; TS-N.
 DR Pfam; PF01849; NAC; 1.
 DR Pfam; PF02094; TS-N; 1.

SQ SEQUENCE 2187 AA; 220599 MW; 003646AA84DEBD CRC64;
 Query Match 15.1%; Score 776; DB 11; Length 2187;
 Best Local Similarity 27.7%; Pred. No. 8,6e-42;
 Matches 288; Conservative 122; Mismatches 389; Indels 240; Gaps 49;
 QY 34 VTTPEDTITQHNKVSPTKITTAKPINRPSLSPNSDTSKE-----TSLTVKETT 84
 Db 806 VQSPKVPDIPMSDVTPTSPKTSATAV--PK-----DTSATLSLKSVPATVSLSPKAV 857
 QY 85 VETKEITTTNKQSTSDKEKTTSAKETQSIKTSANDLAPTSTVKLAPPTKATTGKA 144
 Db 858 APSNEATIVPEIPTSLKNLAATPRETLATSLPKVTSPS---PQKPKSVSLGCA 913
 QY 145 LTT-----PKE-----PTPTTKEBAPT-----PKEPTPT 171
 Db 914 MTSKKATEIAASKDVSSQPEVPLLOHVPPTSPSPVSDTLSCALTSPPKGP--PAT 972
 QY 172 IKSAPTTTKE--PAPTTKSAPTTTKE-----EPAPTTEKAPATTKEBAPT 221
 Db 973 LAETPTTKEKPKAASKKTATPSPGVAVALPEIPCSKKAPKTAAPKESATSSSKR 1032
 QY 222 APTT--TKSAPT-----TPKEBAPTTEK--PAPTTEBAPT 257
 Db 1033 APKTVASKEIPIKGVAVLEISLPLKETSASATPGEKSSASPPKRSPTAGPKE--TP 1089
 QY 258 KEPTTTEKAPATTKEBAPTTEKAPATTKEBAPTTEKAPATTKEBAPT 311
 Db 1090 GGVAVNPEISLPEKPTQONATPNESLAASSQKSPKTSVKEPTPGVAVAMPLEIPSAP 1149
 QY 312 TKEPSPTTPK-----EPAPTTKSAPTTTKEBAPT--KSAPTTK-----EPSP 357
 Db 1150 QKAPTAAPKQIPPEEDAVTILAGSPLSPKKASATAPEKAPKAPVAVSGLSISP 1209
 QY 358 TKEBAPTTEKAPATTKEBAPT--PKE-----PATTEKAPAP--T 396
 Db 1210 KTSKTAAPKENSATLPPKRSKPTAAPKEPTATSEGVAVPSELSPPPTPAKGPV 1269
 QY 397 TTKKAPAPAPKAPATTKEBAPTTEKAPATTKEBAPTTEKAPATTKEBAPT 455
 Db 1270 LTPKAPNLAAP--SPASKKVAPKTAAPETSPTP-----SPOKIPKVAAGKESAVPPS 1322
 QY 456 EPTPTT-----PEBAPTTPKAAANTKEBAPTTPK--PAPTTEKAPATTKEBAPT 506
 Db 1323 KTKPKTAAPKETSAPSEGVAVLEIPSPKAPKTAAPKETAAPVPS--PBQATTAPOIP 1380
 QY 507 PTTTGTATTTKEBAPT-----KKPAKELAPTTTKEPTS-- 544
 Db 1381 PSERKSSKAGSKE--TPTTSPGVTAAAPLEIPSSKTSKMAKSELVLPSSKLSOT 1439
 QY 545 ---TTSKPAATTPKGTAPTTKEBAPTTPK--PAPTTPKGTAPTTKEBAPTTPKPA 599
 Db 1440 VGRKESLGAVALPEIPSHKKAAPKTVDPQVLTLSPK--DAPTTLAE--SPSSPK--A 1496
 QY 600 PKELAPTTTKGPTSTSDKAPATTKEBAPTTPKAPATTKEBAPTTPKAPATTPEP--PT 654
 Db 1497 PKTAAPPSER--VTVPEKPA--TPQASGTTASKVPEPEVQEVAVSSKEPTVTAAPV 1554
 QY 655 TSEVSTPTT-----EPTTIHKSPESTPELSAEPPTKALENSPKE--PGVPTTKTPAA 707
 Db 1555 KNSSHKTSKTELEKAPATLPPSPTKSPKIPSSKAPRT--SAPKEPASPISIK--PVT 1611
 QY 708 TKPEMTTAKDKTTERDLTTEPTTTAAKMKETATTEKTESKIATTTQVSTTQ 767
 Db 1612 T--SLAQTAAPSLQKAPSTTIPKENTLAAPV---LPVSSKSPAPAPARSASLSPTAAP 1665
 QY 768 DTTPKTTTLKTTTAPKVTATTKITLITEINMKPEETAKPKDRATNSKATTPKQKPK 827
 Db 1666 QTAPEKATTTIPSKAAATPTPIETASTASLBAPKEIJE-----TSVAVLMSPP----- 1716
 QY 828 APKKPTSKKPTMPVRKPKTPTPRKMTSTMPLEINPTSLAEAMLOTTTTPNOTPNSK 887

DB 1717 -PKKASSSKRASTLP-----ATLPSIKKEASVLS-----PTATSGK 1752
QY 888 LVEVNPKEADGACGCTP 906
DB 1753 DSHISPV-DACSTGCTTP 1770
RESULT 13
Q9VEL9
AC Q9VEL9 PRELIMINARY; PRT: 2112 AA.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG4090 PROTEIN.
GN CG4090.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Phryganeae; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wen K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell R.M., Besu A.V., Baxendale J., Bayraktaroglu L., Beasly E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burdits K.C., Busan D.A., Butler H., Cadeau E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harits N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Houtlin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegan C.,
RA Jalali M., Kalush F., Karp G.H., Ke Z., Kesterson J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Nelson S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiter K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimons I., Stimpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhou Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003716; AAF55402.1; -
DR FlyBase: FBgn0038492; CG4090.
DR InterPro: IPR002557; Chitin binding.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF01607; Chitin_bind_2: 10.
DR SMART: SM00494; Chetb2: 10.
DR PROSITE: PS00022; EGF_1: 1.
SQ SEQUENCE 2112 AA; 219547 MW; B91018E551A5D36 CRC64;

Query Match 14.3%; Score 739; DB 5; Length 2112;
Best Local Similarity 27.2%; Pred. No. 1.9e-39;
Matches 281; Conservative 98; Mismatches 339; Indels 316; Gaps 44;
QY 4 NKKNRKKKKRPPKPPVYDDEGSLDNGDFKVTTPDSTTHQNKVSTSPKITTAKPINRPP 63
DB 1023 NHPDQIQTKPLKCKKRVYSGGSSN-----STSNSSSSNNSSGSSSSSS----- 1070
QY 64 SLPPNSDTSEKNSLIVNKKETTV---ETKETTNNKQTS---TDCKEKTSAKKEQST-- 114
DB 1071 SSSNSGSSSSNNSSGSSNSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSSSSS 1130
QY 115 -----ETSKADLAPRTSKVLAKPPKAETTYKCP----- 143
DB 1131 NNNNGSSSSSSSSSSSTSKKPNPSECKVNGOFTIGDRSDCAKFRVNDNGGNMVPF 1190
QY 144 -----ALTPKREPTPTTKREPASTPKKEPTPTTKASPTT 178
DB 1191 SCGPGVMDAQACQAHMAVAKECGIAPPTTSTPTSK-PTTASTSRSDQTSRPTG 1249
QY 179 PKEPATTTKSAPTTKPEAPTTTKEPAATTKPEAPTT----- 217
DB 1250 P-----PTTARPYTARPTTSSPTTASSQTTSPVQTAPNTDCKRSEGMADPNCKRYR 1305
QY 218 -----TKRP-----APT-----TKSAPTTKPEAPTTKKRAPTT 248
DB 1306 CVRNKKGFTSTIPFCGAGTWDODLQTCNNHNNCSGTGSTETPKPC--EPATNGTTA 1363
QY 249 PKEPATTKPEPTPTTKPEAPTTKEPAATTKPEAPPAKPPATTKPEAPTTKEPA 308
DB 1364 TSTSTTTP--PTTTDLPTSTTGLP-PTTTTELPTT-----TTTDLPTTTRILPP 1413
QY 309 PTTTEPSTTPKPEAPTTKSAPTTKKEPAATTKSAPTTK--EPPSTTKPEAP-- 363
DB 1414 TTTTSLPPTTTGTPPTTTGAOPTTTTLLSETETSTVTSPESTVQPPSTTKPLPAG 1473
QY 364 -----TTPK-----TTPK-----EPAP---TT 373
DB 1474 TECHGEGMADPEDCRKYRCINMGASRKNFTCPGKGNEEVQCDVYENIPRCSKL 1533
QY 374 PKRAPPTP-----KEPAPTTKE--PAPTTKPPAPTAKEPAPTTKEPAPTTK 428
DB 1534 PAEPPTTTPSEESKDPGCTTPQSTDEPTTVKPI-TKPT-EPSTERKAPPTTQYPRKPT 1591
QY 429 TPKEAPTTPEKPAPTPEELAPTTPEEPTTPEEAPTTPKAAANTPEKAPTTKE 468
DB 1592 TTEE--PEKPKPI-TTEYPOKPTTEEPPEKPKPT-TTEYPOKPTTEEPPEKPK 1647
QY 489 PAPTT-PEKAPTTKEPA-----PTTPK--GTAPTLKEPAPTTPKPAKELAPTTK 540
DB 1648 PTTTEYPOKPTTEEPPTTTSIPGVNPTTTSVGVNPTTPIPIVETTTSTPGK---PTTTC 1704
QY 541 EPSTTSDKAPPTPKGTAPTTKEPAPTTKE--PAPTTKGTAPTTLKEPAPTTK 597
DB 1705 EPITTT-LSTTTDAIQEPTTSKRPPTTTSPESSSTEGSVTTILOPQPNYCSSE 1763
QY 598 ---PAPKELA-----PTTCKPTSTSDKAPPTTKEPAPTTKEPAPTTKPPAPTT 648
DB 1764 GFEPDPEDCGRYRCVDAANKGYAVAFCGGTYWDSTETCNADVOVSGN----- 1816
QY 649 ETTPPTTSEVSTPTTKPEPTTHKSPDESSTPELSAEPTRKALNSKREPVPTTKPAAT 708
DB 1817 ---CSSGQTTTGGTTTPEPTTSTSSGKP-----ETTSKAPEN-----TTTWA-- 1857
QY 709 KPEPTTAKDKTERDLRTPEPTTAPKMTKETATTEKTTSTKTTATTTVOVSTTTOD 768
DB 1858 -PETTT-----SSPETTTT---VASETTTTSGT-----TTTATPEPTTKP 1895
QY 769 TTPPKTTTLTKTLADKVTTKTTTTEIMNKPEETAKPDRAATSKATTPKPKPTKA 828
DB 1896 PKP-----ETTTIAGEESTSKSPTTT-----STAPSTINTSA 1928
QY 829 PKKPTSKKKRKTMP 842

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:13:35 ; Search time 153.5 Seconds

(Without alignments)
896.692 Million cell updates/sec

Title: US-09-556-246-1_COPY_200_1140

Perfect score: 5011

Sequence: 1 VKDNKKRFTKKKPPKPPV.....DMDYLRVNOGIINPMLS 941

Scoring table: BLOSUM62

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organellar:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5011	100.0	1404	4	Q92954
2	5002	99.8	1404	4	Q92954
3	1904	38.0	1054	11	Q92954
4	967	18.3	1079	5	Q92954
5	936.5	18.7	1049	5	Q92954
6	865.5	17.3	1795	5	Q92954
7	853	17.0	1489	10	Q92954
8	806.5	16.1	1315	10	Q92954
9	802	16.0	1274	5	Q92954
10	800.5	16.0	1480	10	Q92954
11	784	15.6	1188	10	Q92954
12	776	15.5	2187	11	Q92954
13	739	14.7	2112	5	Q92954
14	715	14.3	763	2	Q92954
15	707.5	14.1	555	10	Q92954
16	671	13.4	1151	13	Q92954
17	667.5	13.3	4880	11	Q92954
18	667.5	13.3	5085	11	Q92954
19	660.5	13.2	6677	5	Q92954

20	659.5	13.2	1229	5	Q94185
21	654	13.1	2284	5	Q94185
22	653.5	13.0	1514	5	Q94185
23	647	12.9	4833	11	Q94185
24	647	12.9	5038	11	Q94185
25	641.5	12.8	2089	4	Q94185
26	633	12.6	489	10	Q94185
27	632	12.6	761	10	Q94185
28	632	12.6	7962	4	Q94185
29	629	12.6	3507	5	Q94185
30	629	12.6	6632	5	Q94185
31	626.5	12.5	990	13	Q94185
32	624.5	12.5	1612	5	Q94185
33	624.5	12.5	2768	5	Q94185
34	622.5	12.4	839	2	Q94185
35	622	12.4	6642	5	Q94185
36	620.5	12.4	971	5	Q94185
37	617	12.3	3570	5	Q94185
38	607.5	12.1	801	5	Q94185
39	607.5	12.1	924	12	Q94185
40	606.5	12.1	2344	5	Q94185
41	605	12.1	379	5	Q94185
42	600.5	12.0	1893	5	Q94185
43	598.5	11.9	409	10	Q94185
44	577	11.5	956	10	Q94185
45	573	11.4	2601	4	Q94185

ALIGNMENTS

RESULT	ID	SEQUENCE FROM N.A.	PRELIMINARY:	PRT:	1404 AA.
Q92954	Q92954	01-FEB-1997 (TREMBLrel. 02, Created)			
Q92954	Q92954	01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
Q92954	Q92954	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
Q92954	Q92954	MEGAKARYOCYTE STIMULATING FACTOR.			
Q92954	Q92954	OS Homo sapiens (Human)			
Q92954	Q92954	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
Q92954	Q92954	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
Q92954	Q92954	NCBI_TaxID=9606;			
Q92954	Q92954	[1]			
Q92954	Q92954	SEQUENCE FROM N.A.			
Q92954	Q92954	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
Q92954	Q92954	Kellerer K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,			
Q92954	Q92954	Frederick C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
Q92954	Q92954	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
Q92954	Q92954	Blood 78:279-279(1991).			
Q92954	Q92954	[2]			
Q92954	Q92954	SEQUENCE FROM N.A.			
Q92954	Q92954	Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,			
Q92954	Q92954	Fitzgerald M., Scallitro J., Kellerer K., Preissner K., Kriz R.,			
Q92954	Q92954	Jacobs K., Turner K.,			
Q92954	Q92954	(In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,			
Q92954	Q92954	Mosher D.F. (eds.);			
Q92954	Q92954	Biology of vitronectins and their receptors., pp.45-52,			
Q92954	Q92954	Elsevier Science Publishers B.V. (1993).			
Q92954	Q92954	[3]			
Q92954	Q92954	SEQUENCE FROM N.A.			
Q92954	Q92954	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
Q92954	Q92954	Kellerer K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,			
Q92954	Q92954	Frederick C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
Q92954	Q92954	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
Q92954	Q92954	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.			
Q92954	Q92954	EMBL, U70136; AAB09089.1;			
Q92954	Q92954	InterPro: IPR000585; Hemopexin.			
Q92954	Q92954	InterPro: IPR001212; Somatomedin_B.			
Q92954	Q92954	InterPro: IPR002400; GF_cys_knot.			
Q92954	Q92954	Pfam: PF00045; hemopexin; 2.			
Q92954	Q92954	Pfam: PF01033; Somatomedin_B; 2.			
Q92954	Q92954	PRINTS: PR00438; GFCYSKNOT.			

DR PRINTS: PS00022; SOMATOMEDIN.
 DR PROSITE: PS00024; HEMOPEXIN; UNKNOWN_1.
 DR PROSITE: PS00524; SOMATOMEDIN_B; 2.
 DR SMART: SM00120; HX; 2.
 DR SMART: SM00201; SO; 2.
 SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35FAF6 CRC64;

Query Match 100.0%; Score 5011; DB 4; Length 1404;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKDNKKNRKTKKPPKPPVVDAGSGLDNGDKVTPDSTIOHKNVSTSPKITTAKPIN 60
 DB 200 VKDNKKNRKTKKPPKPPVVDAGSGLDNGDKVTPDSTIOHKNVSTSPKITTAKPIN 259
 QY 61 PRPSLPNSDTSKETSILYNKETTVEKETTNNKOTSDGKEKTSAKETOSIEKTSK 120
 DB 260 PRPSLPNSDTSKETSILYNKETTVEKETTNNKOTSDGKEKTSAKETOSIEKTSK 319
 QY 121 DLAPTSKVLAKPTPKAETTTKGPALTPPKPPPTPKKEPASTTPKBPPTTKSAPTTK 180
 DB 320 DLAPTSKVLAKPTPKAETTTKGPALTPPKPPPTPKKEPASTTPKBPPTTKSAPTTK 379
 QY 181 EPAPTTKSAPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 240
 DB 380 EPAPTTKSAPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 439
 QY 241 PKRAPTTPEPAPTTKEPTPTPKKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPA 300
 DB 440 PKRAPTTPEPAPTTKEPTPTPKKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPA 499
 QY 301 PTPPKKEPAPTTKEPSTTPPKKEPAPTTTKSAPTTTKEPAPTTTKSAPTTTKEPSTTTTKE 360
 DB 500 PTPPKKEPAPTTKEPSTTPPKKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPSTTTTKE 559
 QY 361 PAPTTPKEPAPTTPKKAPPTPKKEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 420
 DB 560 PAPTTPKEPAPTTPKKAPPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 619
 QY 421 TPKKLAPTTPEKLAPTTPEKAPTTPEELAPTTPEEBPTPTPEEPAPTTTKEAANPTPK 480
 DB 620 TPKKLAPTTPEKLAPTTPEKAPTTPEELAPTTPEEBPTPTPEEPAPTTTKEAANPTPK 679
 QY 481 PAPTTPKEPAPTTPKKAPPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 540
 DB 680 PAPTTPKEPAPTTPKKAPPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 739
 QY 541 EPTSTSDKAPPTPKGTAPTTPKKEPAPTTKEPAPTTPKGTAPTTTKEPAPTTPKKAP 600
 DB 740 EPTSTSDKAPPTPKGTAPTTPKKEPAPTTKEPAPTTPKGTAPTTTKEPAPTTPKKAP 799
 QY 601 KELAPTTTKEPSTTSOKPAPTTPKETAPTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 660
 DB 800 KELAPTTTKEPSTTSOKPAPTTPKETAPTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 859
 QY 661 PTTTKEPPTTIHKSDESTPELSAPETPKALENSPEKPGVPTTKTTPAATKEMTTTAAKDT 720
 DB 860 PTTTKEPPTTIHKSDESTPELSAPETPKALENSPEKPGVPTTKTTPAATKEMTTTAAKDT 919
 QY 721 TERDLRTPTETTTAAPKMKETATTTETKTESKTTATTTQVSTTTQDTPPFKTTTAKTT 780
 DB 920 TERDLRTPTETTTAAPKMKETATTTETKTESKTTATTTQVSTTTQDTPPFKTTTAKTT 979
 QY 781 TLAKVTTTAKTTTTELMNKEETAPKPKDRAATNSKATTPPKOKTAPKAPKPTSTKKPKT 840
 DB 980 TLAKVTTTAKTTTTELMNKEETAPKPKDRAATNSKATTPPKOKTAPKAPKPTSTKKPKT 1039
 QY 841 MPKRRKKTPTPTPKKMTSTIMPELNPSTKRIAEAMLQTTTRPQOTPNKSLVEVNPKSEADAG 900
 DB 1040 MPKRRKKTPTPTPKKMTSTIMPELNPSTKRIAEAMLQTTTRPQOTPNKSLVEVNPKSEADAG 1099
 QY 901 AEGETPHMLRPHVFMPEVTPMDYLRVVPNOGIINMLS 941

DB 1100 AEGETPHMLRPHVFMPEVTPMDYLRVVPNOGIINMLS 1140

RESULT 2 PRELIMINARY; PRT: 1404 AA.
 ID Q9BX49
 AC Q9BX49
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
 DE BG174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR).
 GN BG174L6.2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wray P.;
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL13553; CAC36090.1; ..
 SQ SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEB5 CRC64;

Query Match 99.8%; Score 5002; DB 4; Length 1404;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 939; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VKDNKKNRKTKKPPKPPVVDAGSGLDNGDKVTPDSTIOHKNVSTSPKITTAKPIN 60
 DB 200 VKDNKKNRKTKKPPKPPVVDAGSGLDNGDKVTPDSTIOHKNVSTSPKITTAKPIN 259
 QY 61 PRPSLPNSDTSKETSILYNKETTVEKETTNNKOTSDGKEKTSAKETOSIEKTSK 120
 DB 260 PRPSLPNSDTSKETSILYNKETTVEKETTNNKOTSDGKEKTSAKETOSIEKTSK 319
 QY 121 DLAPTSKVLAKPTPKAETTTKGPALTPPKPPPTPKKEPASTTPKBPPTTKSAPTTK 180
 DB 320 DLAPTSKVLAKPTPKAETTTKGPALTPPKPPPTPKKEPASTTPKBPPTTKSAPTTK 379
 QY 181 EPAPTTKSAPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 240
 DB 380 EPAPTTKSAPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 439
 QY 241 PKRAPTTPEPAPTTKEPTPTPKKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPA 300
 DB 440 PKRAPTTPEPAPTTKEPTPTPKKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPA 499
 QY 301 PTPPKKEPAPTTKEPSTTPPKKEPAPTTTKSAPTTTKEPAPTTTKSAPTTTKEPSTTTTKE 360
 DB 500 PTPPKKEPAPTTKEPSTTPPKKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPSTTTTKE 559
 QY 361 PAPTTPKEPAPTTPKKAPPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 420
 DB 560 PAPTTPKEPAPTTPKKAPPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 619
 QY 541 EPTSTSDKAPPTPKGTAPTTPKKEPAPTTKEPAPTTPKGTAPTTTKEPAPTTPKKAP 600
 DB 680 PAPTTPKEPAPTTPKKAPPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 739
 QY 421 TPKKLAPTTPEKLAPTTPEKAPTTPEELAPTTPEEBPTPTPEEPAPTTTKEAANPTPK 480
 DB 620 TPKKLAPTTPEKLAPTTPEKAPTTPEELAPTTPEEBPTPTPEEPAPTTTKEAANPTPK 679
 QY 481 PAPTTPKEPAPTTPKKAPPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 540
 DB 680 PAPTTPKEPAPTTPKKAPPTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 739
 QY 541 EPTSTSDKAPPTPKGTAPTTPKKEPAPTTKEPAPTTPKGTAPTTTKEPAPTTPKKAP 600
 DB 740 EPTSTSDKAPPTPKGTAPTTPKKEPAPTTKEPAPTTPKGTAPTTTKEPAPTTPKKAP 799
 QY 601 KELAPTTTKEPSTTSOKPAPTTPKETAPTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 660
 DB 800 KELAPTTTKEPSTTSOKPAPTTPKETAPTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 859

	Query Match	38.0%;	Score 1904;	DB 11;	Length 1054;
	Best Local Similarity	44.4%;	Pred. No. 3.56-113;		
	Matches 423;	Conservative 33;	Mismatches 128;	Indels 368;	Gaps 20;
OY	1 VKDKKKRRTKPKPPVVDAGSGDLNDGPKVH--TPDSTQHNVKSPTIKTAKP	58			
Db	: : : : : : : : :				
OY	195 VKDKMKTKPKKPPEPAVDEGGSLNGEKLPRPPDPPTPHSKVAISRTTKNAKP	254			
Db	: : : : : : : :				
OY	59 INRPRLPSPDSQSKETSILTVNKEETVVETKETTTNNOSTGDGEKTTSKAKOSIEKTS	118			
Db	: : : : : : :				
OY	255 VTRPSLAPSEISEKASIASLNKEITVEKETAATNKOSA-SKKKTSYKVEFRSAEKTS	313			
OY	119 AKDLAPTISKVLAKPTTKEATTGPGALTPRKEPPTTPKEASDTPKEPIPTTIKKAAPT	178			
Db	:: :				
	314 DKDV-----NSAPTT-----EPSTSTR-----NSAPTT 331				

[illegible]

```

RL Science 282:2012-2018 (1998).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Cotton M.;
RT "The sequence of C. elegans cosmid Y51B1A.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; Acc06797; AAP60743.1; -.
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01217; PRICHEXTNSN.
SQ SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;

Query Match 19.3%; Score 967; DB 5; Length 1079;
Best Local Similarity 32.3%; Pred. No. 8,4e-54;
Matches 294; Conservative 114; Mismatches 377; Indels 124; Gaps 33;

OY 35 TTPPTSTTHNKKVSTSP-KITTAAPINRPSLPPNSDTS--KETSLLVKKETVEKETT 91
DB 102 TAPETSTSEAPSSSTTPVQTTTAPETTTSTEPSSTSPVQTTTAPETTTSEAPSS 161
OY 92 TTNKQTSIDGKEKTSKMETQSIKTSKDLAPTSKVLAKPTPKAETTKGP--ALTTPK 149
DB 162 TTPVQTT-----TTPAETTTSTEPSSTSS--TSPVQTTTAPETTTSTEPSSTTPV 211
OY 150 EPTPTTPKEPASTTPKEPTPTTIKSAPTTPEKAPATTTSKAPTTPEKAPATT 209
DB 212 QTTTITTAETTTSTEPSSTTPVQTTTAPETTTSTEPSSTSPVQTTTAPETTTSE 268
OY 210 PKREPATTTKKAPATTTSKAPTTPEKAPATTTPPKAPATTPEKAPATTPEKPA 269
DB 269 STEP-PSSSTTPVQTTTAPETTTSTEPSSTTPVQTTTAPETTTSTEP-PSSSTTPV 326
OY 270 PTKKAPATTPEKAPATTTPKAPATTPEKAPATTPEK-----APTTTKEP 315
DB 327 QTTTITTAETTTSTEPSSTTPVQTTTAPETTTSTEPSSTTPVQTTTAPETTTSE 386
OY 316 SPTTTPKEPATTTKSAP--TTTKEPAPTITTSKAPTTPEKAPATTPEKAPATT 372
DB 387 PPSSTTPVQTTTAPETTTSTEPSSTTPVQTTTAPETTTSTEPSSTTPVQTTT 446
OY 373 PTKKAPATTPEKAPATTTPKAPATTTPKAPATTPEK--APTTPEK-----TAP----- 419
DB 447 T-----APETTTSTEPSSTTPVQTTTAPETTTSTEPSSTTPVQTTTAPETTTSTEP 502
OY 420 ----TTPKRLPTTPEKLAAPTPEKAP--TTPEELAPTPEEPTPTTPEE----PAPT 468
DB 503 PSSSTTPVQTTT-----APETTTSTEPSSTTPVQTTTAPETTTSTEPSSTTPVQTT 558
OY 469 TPKAAPNTPKPE--APTTPEKAPATTPEKAPATTPEKAPATTTPKGTATTTKEPAPT 526
DB 559 TTTTAAETTTSTEPSSTTPVQTTTAPETTTSTEPSSTSS--TTPVQTTTITAPETTTSTEP 617
OY 527 KKPAPELAATTTKEPTSTTSDDK--APTTPKGAPTTPKEAPATTPEK--APTTPKGTA 583
DB 618 PSSSTTPVQTTTAPETTTSTEPSSTTPVQTTTAPETTTSTEPSSTTPVQTTTAPETTTSTEP 674
OY 584 PTLKAPATTTPKAPKAPKELAPTTTGTSTSTSDK--APTTPEKAPATTPEKAPATTPK 642
DB 675 TTTTAPETTTSTEPSSTTPVQTTTAPETTTSTEPSSTSSSPVQTTTITAPETTTSTEP-- 732
OY 643 PAPTTPPTTSEVSTPTTKEPTTIHKSPEDESTPELSAFT--PKALENSKPEKPV 699
DB 733 --PSSSTTPVQTTTITAPETTTSTEPSSTTPVQTTTITAPETTTSTEPSSTNTP--V 786
OY 700 PTTKPAATPEMTTAKKDTTERDLRT-----PETTTAPKMTKATTTKTEKTSKIT 755
DB 787 QTTT-----TAPETTTSTEPSSTTPVQTTTITAPETTTSTEPSSTNTPVQTTTITAPETTT 843

```

```

OY 756 ATTQVSTTTTQDTTPPKITTLTKTTTAPKVTYTK-----KITTTIEIMKKPEETA-KPK 809
DB 844 STEPPSSSTSPVQTT-----TTTAPETTTSTEPSSTTPVQTTTITAPETTTSTEP 894
OY 810 DRATNSKATTTKPKPKPTAPKPKPTSTKKPKTPMVRPKPTTPPRKMTSTMPELNPTSRI 869
DB 895 SSSTTPVQTTTITAPETTTSTEPSSTTPVQTTTITAPETTTSTEPSSTTP----- 946
OY 870 AEAMLOTTT 878
DB 947 -----VQTTT 951

RESULT 5
ID 0917S1 PRELIMINARY; PRT; 1049 AA.
AC 0917S1
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE CG52228 PROTEIN.
GN CG52228.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Ananthias A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sultoni G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Chape M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Adayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Boishakov S.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butcher H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostel A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Heiman T.J., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jbail M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshirel A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Schaefer F., Shen H.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shie B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
DR EMBL; AE003495; AAG22353.1; -.
DR FlyBase; FBgn0030561; CG52228.
SQ SEQUENCE 1049 AA; 107278 MW; 954DD629E7619671 CRC64;

```


DR FlyBase: FBgn0025390; EG:5667.1.
 DR InterPro: IPR002557; Chitin_binding.
 DR InterPro: IPR002965; P_titch_extensn.
 DR Pfam: PF01607; Chitin_bind_2.
 DR PRINTS: PR01217; PRICHTEXTENS.
 DR SMART: SM00494; ChitBD2.2.
 DR SPOUNCE 1795 AA; 19446 MW; 07F10C129BD9557B CRC64;

Query Match 17.3%; Score 865.5; DB 5; Length 1795;
 Best Local Similarity 28.3%; Pred. No. 3.5e-47;
 Matches 276; Conservative 129; Mismatches 447; Indels 115; Gaps 28;

DB 6 KNRKKKPPVVDAGSGDNGDFKVTTPDTSTOHNKVSTSPKTTAKPINDPRL 65
 431 KNDVTEAPEIKSPL-----KQHLSENVILPEPTTT-TTTTTKPVLLCPRLSPDDT 483

DB 66 PPNSDT-----SKETSLTVNKEVETVEKETTNNKQSTSDGKEKTSAGEQST 114
 484 PKPSTTAVTKSTPKISTEQHSITTKAKTTTKRPTTVTEKTSATEKPRPTVVTTTOK 543

DB 115 EKTSAKLAPTSKVLAKPTPKAETTKGPAITTPKEPTTPPKPASTTPKEPTTIKS 174
 544 RSTTHNTSPDKTTIRSTLSPKTTTPSTTTSTTPST-----TTSSTTPST--- 594

DB 175 APPTPKPAPPTTKSAPPTPKPAPPTTK-BPAPPTPKPAPPTTKPAPPTTKSAPPTP 233
 595 --TPSTTTPSTTTTVKSTHRPRTSQTTASTTTK-----TTTSPKTTKTIDTST 648

DB 234 KEPAPPTTKKAPPTPKPAPPTTKREPPTTPPKPAPPTTKPAPPTTKPAPPTPKPAP 293
 649 TSKLSTTQKTTTHHKTAATSTKPKTTTEKTSIVSTTKTSSTESPKPTSTGKPT 708

DB 294 TPKPAPPTTKPAPPTTKPAPPTTKPAPPTTKSAPPTTKPAPPTTKSAPPTTKP 353
 709 TTPKSTRTPTTTKVTITTTQITTTPLRSSSTETSTQPTTTTPQPTTTTLVTPKTS 768

DB 354 SPPTTKPAPPTTKPAPPTTK-----KPAPTTKKEAP-----TPKKE 392
 769 TTTTTEKTISSPKPTTTOKTSTAPNTKVALITTEKTPPTQOSTITTTFTKRTITNN 828

DB 393 PAPPTTKKAPAPPTPKPAPPTTK-ETAPPTPKLPTTPPEKLIAPPTPKP-----AP 443
 829 PEPSTSEKRTITTPKSTTTTPKTSVASTKSTISSPKPTKESTENPTNSVTSAL 888

DB 444 TPPELAPPTPEEPPTTPPEPAPPTTKAAANTKEAPPTPKPAPPT--TPKAPAPT 501
 889 TSSSTORATSTSEPTKT--QNTITTTTPKPTTLKSTQEAITSSTQVSTVTTTKATSS 947

DB 502 PKETAPPTTKGAPPTTKPAPPTTKPAPPTTKKELAPTTTKEPT--STTSKAPPTPKGA 559
 948 PLTTTSTEEPNTPKLKTPTTTSVATRTITTTTISSESTETTSQPKSTPTSTT 1007

DB 560 PTPPKPAPPTTKPAPPTTKGAPPTTKPAPPTTKPAPPTTKKAPKELAP--TTTKPPTSTSDK 618
 1008 RTTPKVTIVIVSQNPPTTTSKTSVTI-----TTP-NPSSTORPTTTTTPOTSTIATST 1061

DB 619 PAPPTTKKAPPTTKPAPPTTKPAPPTTKPAPPTTPETPTTSVSTPT--TKP----- 666
 1062 TSTGTTRIPPTTPONSTSTDLITVTRPCPDPSSTSKNTNTACTOQLQOVNLELOS 1121

DB 667 -----PTTIKSPDESTPELSAPTPKALENSPKEGVPPTTKPAATKBEMLTTAKD 718
 1122 POKOEOFTHTHTALGSAANTLGGQEVDPYMDAPSSAEAGQATTTAKAPMTSLAAA 1181

DB 719 KTER---DLRTPEITTAAPKMTKETATTEKTSKATITATTOVTSTTTOPTPEKIT 775
 1182 HLLQKLEHIISTTPSKREAP--TORSSOPSSORSR-GVITAOAARHNLATSKFFIH 1238

DB 776 TTKTT---TLAPKVTTKTITTTTEIMNKEPETAAPKRAATNSKATTPKPKQPTKA-PK 830
 1239 SLRLSTIOQLASTOKRSIRPRTIVTHTTKPED-SEVYDSETSEQVTDDENEVLDTOPR 1297

DB 831 KPTSTKPKTMPRVKPKPTTPPKMTSPMLNPSTRIAEAMLOTTTREPNOTPNSKIVE 890
 1298 AMSSITVAVALPAPVSTTTEREPOK-TSSSP--SPT-----KATSTTTTQILETTTGOLEY 1350

DB 891 VPKPSED 897
 1351 DSSGSSD 1357

RESULT 7
 096449 PRELIMINARY; PRT; 1489 AA.
 ID 096449;
 AC 096449;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, last annotation update)
 DE CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.
 GN CAR90.
 OS Phytophthora infestans (Potato late blight fungus).
 OC Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
 OC Phytophthora
 OC NCBI_TaxId=4787;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=RACE 1-11;
 RA Goenhardt B.;
 RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF061185; AAC72308.1;
 DR SPOUNCE 1489 AA; 164037 MW; 764CC79D1C2F5163 CRC64;

Query Match 17.0%; Score 853; DB 10; Length 1489;
 Best Local Similarity 32.3%; Pred. No. 1.8e-46;
 Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;

DB 9 TKKPPKPPVVDAGSGDNGDFKVTTPDTST-----OHNKVSTSPKTTAKPINDR 62
 338 TPYAPTEKPYDVEETTYVTEESTYAPTKSETNAPRHMVAHLEKCDQEVMTYAPTEET 397

DB 63 PSLP-----PNSDT-----SKETSLTVNKEVETKETT---TNKQSTSD 100
 398 TYAPTEETTYAPTEETTYAPTEETPEPEPTTEETTYAPTEETTYAPTEETTYAP 457

DB 101 GKEKTTAKETOSIEKTSADLAPTSVLAKPAPKAEKTTKGPAITTPKEPTTPKPEA 160
 458 TEETTYAPTEETPEPEPTTEETTYAPTKETTYAP---EETTYASTETTYAPTEETTYA 514

DB 161 STPKPEPTPTTIKSAPPTPKPAPPT--TKSAPPTPKPAPPT--TKPAPPTPKPAPPT 217
 515 EETPEPEPEE-TPAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPEPEEET 573

DB 218 TKPAPPTTYSAP-----TKPAPPTPKKAPPTPKPAPPTPKPAPPTPKPEPTTPKE 267
 574 TYAPTEET--YAPTEETMYAPIETTYAPTEETTYAPAEETPEPEPTTEETTYAPTE 632

DB 268 P-----AP-----TKPAPPTPKP-----APAPKAPAPPTPKPAPPTP 304
 633 PTEETTYASTETTYAPTEETTYAPAEETPEPEPTTEETTYAPTEETTYAPTEETTYA 692

DB 305 KEPAPT--TKKPSPTPKP-----APT--TKSAPT-----TKEP-----A 339
 693 TTYAPTEETTYAPAEETPEPEPTTEETTYAPTEETTYAPAEETMYAPIETTYAGTEETTY 752

DB 340 PT--TKSAPT--TKKPSPTTKP-----APTTPKAPAPPTPKKAPPTPKPAPPT-- 368
 753 PTEATTYAPTEETTYAPAEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYA 812

DB 369 TPKEPAPPTTKKAPAPPTPKAPAPPTPKETADTPPKKLPTTPPEKLAPTTEKAPPTPEE 448
 813 TPPEPEETTYAPTEETPEPEPEE--TYPTTEETTYAPAEETTYAPTEETTYAPTEET 870

DB 449 IAPT--TEPEPTPTPEP-----APT-----TPKAAPNPKEPAPPTPKP----- 489

Db 871 YAPTEETPYEPTETTYAPTEETTYAPTEETTYASTETTYAPTEETTYAPAEPTET 930
QY 490 -----APTPKPEAPPTPKETAPTPKGNAPT--TKAPEAPTPPKPAK--ELAPT--TT 539
Db 931 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEPTETPYEPTETTYAPTEET 990
QY 540 KEPTSTTSOKPAPPTPKGAPPTPKPEAPPTPKPEP-----APTPKPGAPT----- 585
Db 991 YAPTEETMYAPTEETTYAPTEETTYAPAEPTETPYEPTETTYAPTEETTYAPTEETTYAST 1050
QY 586 -----TKAPEAPTPPKP-----APKE--LAPT--TYGAPSTTSOKPAPPTP 624
Db 1051 EETTYAPTEETTYAPAEPTETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1110
QY 625 KETAPTPKPEAPPTPKKAPPTPKPEPTPEPTSEVSTPTTKAPT--TIHKSPESTPELS 682
Db 1111 YAPAEPTPYEPTETTYAPTEETTYAPTEETMYAPTEETTYAPTEETTY--APTEATYAP 1169
QY 683 AEPTPKA--LENSPKERG-----VPTTKTPAATPKPEMTTAKDKTERDLRTTP--BT 732
Db 1170 TEETPYAPTEETTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPTET 1229
QY 733 TAAPMKTEATTTTEKTESKITATTTTQVSTTTQDT-----PKITT--LKT 779
Db 1230 TYAP-----TEETTYEPTETTYAPTEETTYAPTEETTYAPTEETMYAPTEETTYAPTEET 1284
QY 780 TTLAKRVTTTKTTTTEIMNKPEETAKPKDRAT--NSKATTPKP--QKTPAKPKPT-- 833
Db 1285 TYAPTEATYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1344
QY 834 -----STRKKP--KTPBRVKKPTTTPPKMTSTIMELAPTSIAEAMLTTPP--N 881
Db 1345 PAESTSTVSTEKPCNTTEETDEPTDEPTDE--PSDEPTDEPTDEPTDEPTDEPTDCDN 1402
QY 882 QTPNSKLEVNPKSEDAG 899
Db 1403 QGINGIGVENKRYNNAG 1420

RESULT 8
Q9SPMO 8
ID Q9SPMO PRELIMINARY; PRT; 1315 AA.
AC Q9SPMO;
DT 01-MAY-2000 (TREMBLERel. 13, Created)
DT 01-MAY-2000 (TREMBLERel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLERel. 17, Last annotation update)
DE EXTENSIN-LIKE PROTEIN.
CN PEK2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-POLLEN;
RA Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA Hohorst D., Gao M., Showalter A., Bedinger P.A.;
RT "Hollen Extensin-like (Pex) Genes in a Monocot and a Dicot."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AEL59297; AAD55980.1; -
DR InterPro: IPR001611; LRR.
DR InterPro: IPR001998; Xylose_isom.
DR InterPro: IPR002965; P_rich_extensin.
DR InterPro: IPR003592; LRR_out.
DR Pfam: PF00560; LRR_3.
DR PRINTS: PRO1217; PRICHEXTENS.
DR PROSITE: PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR SMART: SM00370; LRR; 3.
SQ SEQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;

Query Match 16.1%; Score 806.5; DB 10; Length 1315;

Best Local Similarity 25.0%; Pred. No. 1,5e-43;
Matches 255; Conservative 86; Mismatches 459; Indels 219; Gaps 33;

QY 13 PTP-----KPVYDEAGSLGNDGKVTTPDTSTTHNKVSTSPKITTAKPINPR 62
Db 412 PTPGGPSSSVGKPPSVG-----KPAAPAMPPIHPDPVSP-----EPL-PE 455
QY 63 PSLRP-----NSDTSKETSLTVNKETTYETKETTJTNKQTSIDGKEKTSKENEQ 112
Db 456 PSVPAPAPMGMPLRSPDAVEIPT-----PPVAKSPPEGTSPASGAPPLAQAOPAS 511
QY 113 SIEKTSKDLAPTSKY-----LAKTPAEETTKGAPALTTPKEPTTPKE 158
Db 512 SPATPVKSSPPAAVVLPPAPKPSAPAVASHPPEAPVSPQGVKSSPPAPVASP 571
QY 159 PASTTPKEPTTITKSAPTTPKEAPPTTSAPTTP--KEAP-----TTKEAPATTP 210
Db 572 PPMKSP--PPPARVASPPMLKSPPRAPVASPPQPLKSPPPVLMSTPVSPPPPV 629
QY 211 KEAPATTKAPATTTTSAPTTPKE-----PAP--TTP--KKAPATPKAPATTPKE- 259
Db 630 VASPPPVKSPPLAPVSSPSPVKLPPLAPGKSTPPPEEKPTPTPVKSSPPEKSL 689
QY 260 PTPTPKEAPATTKAPATTPKEAPATAPKKAATTPKEAPATTPKEAPATTTKESPTT 319
Db 690 PPTLTTSPPQEKPTPTPTSTSKPPPSVETLPPSKSSPPEEPVSSPOARKSSP-- 747
QY 320 KPEAPATTTKSADPTTKAPATTTKSAPTTPKEPSPTTTPKEAPATTPKEAPATTPK 379
Db 748 ---PAPVS--SPPLKSSPPPVESSPTPKSSPLAPVSSPPQVEKTSPPAPVSSP 802
QY 380 TTPKEAPATTPKEAPATTTTKKAPATAPKKEAPATTPKEATPTPKLTPTPEKLAPTPE 439
Db 803 PTPKSSPPLAPVSSPPQVEKTSPPAPVSSPPTPKSSPLAPVSSPPQVEKTSPPAPV 862
QY 440 KPAATTPPEELAPTPPEEPTTPPEEAPATTPKKAANTTPKEAPATTPKAPATTPKEAP 499
Db 863 SSPLEPKSSPPSSVSPPTTKSSPPAPVSSPMTKSSPPAHVSSPPEAEKSSP 922
QY 500 TTPKETAPTPKTA-----PTTLKEAPATTPKPAKELAPTTKEPTSTSD 548
Db 923 LAPISSPSESPSPSPVWEKTSPPATVSSPPTPKSSPP--APVSSPPVVKSP 979
QY 549 KPAPT-----TPKGATTPP-----KEAPATTPKEAPATTPK--GAPLTTLKEAP 593
Db 980 PPAVSSPPTPKPLPPAPVSSPPPVKSSPPPTPVSSPPTPKPLPPPTPVSSPPT 1039
QY 594 TPKKAPAKELAPTTKGPSTTSDDKAPATP--PKETAPTPK-----EPAPTPKKAPAT 647
Db 1040 XKRLLPP--APVSSPPPVKSSPPAPVSLPPTPKRKSPPRTKVSRRPVVKKCCPPTL 1096
QY 648 PEPPTTSEVSTPTTTKPTTIHKSPESTPELSAEPTPKALENSPKPEGVPTTKTPAA 707
Db 1097 VSSPPPAKSLPPTPVSSP-----PPEVSSPPPTPVSSPPAPKSSPPTPV 1146
QY 708 TPPEMTTAKDKTTERDLATTP-----ETTYAAPMTKETATTTKTESKITATTTQVT 762
Db 1147 SPP-----ELKSSPPAPVSSPSPASAPKSSPPAPVSLPPEVVKSSPPAPIS 1194
QY 763 STTTODTTPFKLTTLKTTTLAPKVTTKTTTTEIMNKPEETAKPKDRANSKATTPKP 822
Db 1195 SPPPAKSP-----PPPA-----PMSLLEPVKSSPPAPVSSP 1228
QY 823 QKTPAKPKPTSTKKKTPMPVRKKTTPTPRKMTSTIMELNPTSRILAEAMLQTTTPNQ 882
Db 1229 PPPKSSPPPAVSSPPAP--VAPSLPPAPVSSPPAPV-----TSAP-- 1271
QY 883 TPNSKLEVNPKSEDAGAEGETPHMLLRPHVFMPEVTTPDDVTLPRVNOGIINPMIS 941
Db 1272 -----PKKEEDSTA-----PRAELPPSPENDIILPIMA 1301

RESULT 9

Q20007
ID Q20007 PRELIMINARY: PRT: 1274 AA.
AC Q20007;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE COSMID F35A5.
GN F35A5.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifen L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton J., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RP elegans".
RL Nature 368:32-38(1994).
RN
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Leimbach D.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U46675; AAB52641.1;
SQ SEQUENCE 1274 AA; 138065 MW; 07F6BD292C1799F CRC64;

Query Match 16.0%; Score 802; DB 5; Length 1274;
Best Local Similarity 28.0%; Pred. No. 2,7e-43;
Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

QY 6 KNRTRK-KP-----TP-----KPYVDKAGSLDNGD-----FKYTTPTSTQHMKVSTS 50
DB 274 KNPTRKMPWDETPVEVEKPEPVKAPVLRKRDPAKADPSKAPKAVES 333
QY 51 PKTTAKPI-NRPSLPKNSDTSKETSITVNEKETTVEKETTNNKQSTDSKEKTTSAK 109
DB 334 SPVVPPTPVKNPVKTKKPPWEVDDEAEVKKPSAPEKTPVLKRRKEPPSSTPSSDPS 393
QY 110 ETOSIEKTSKADLAPTSKVLAKPTPKAE-----TTTKGPA-----LTTT 148
DB 394 PKAAAVKPRDSSPKKATPLADPKAQEVPTPVKNPVKPKPPWEVDDEPVEVKOP 453
QY 149 KEPTPTTP-----KEPASTTPKEP-----TP-TTKSAPTPKPEPAPTTSKAPTTPKEPAP 199
DB 454 EAPAKKTPVLKRRKPEPAKDTAKATGTETPEPKDPVKPRDSSPKVAAKPSAARAP 512
QY 200 TTKTEPA-----PTTPKEPAPTTPKEPAPT-----TTKSAPTTP 233
DB 513 TPVKNVKKRRPPWDETPADVDVSKPTDAKTPSLAKKDPAPAKBSLKPADTKAPAP 572
QY 234 KEP-----APTTPKPAPTTPKE-----PAFTTKEPTPTTPKEPAPTTPKEPAPTTPK-EP 283

DB 573 RDPSPKVAAPTAPEKKTIVLAKKEPAGPADSKTEPEKSKRDPDSPKKAIVAKVPEKTEV 632
QY 284 APTAKKPPAP-----TTPKEPAPTTPKEPAPTTPKEP-----SPTTPKEPAPTTPKAP 333
DB 633 ABAVAKKEPKSPKDTAPKKAEPNSVPV-PTPVKNPVKWKMPWEDDAPAPVSLPE 691
QY 334 TTKEPAPTTPKAPITTP-KESPPTTKKEPAPTTPK--EPAPTTPKPPAP--TTPKEPAPT 388
DB 692 PEKK--TPVLAKKAPTKPDSEAADPVSGPSSKDPKLAKAAVYKPRDSSPKAAVYIKAPK 750
QY 389 TTKEPAPTTPKAPTA--PKEPAPTTPKEPAPTTPKKTLPPTP-----EKLAPTTPKPA 442
DB 751 T--EVPNAVVKKEPVAKSRDPSKKAK--AEVNSP--VVPPTPVKNPVKWKMPWEDDADA 805
QY 443 PTTPEELAPTTPEEP-----TTPPEEPAPTTPKAAANTKREPAPTTPKEPAPTTPK 496
DB 806 PAEPVAVPEPEKKTPLAKTTPVKPRDSSPKKAVPAKSTKTADAPVVKKEPEVSKPE 865
QY 497 PAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPT--TTPKEPSTSDKPAPTTP 555
DB 866 PSFKKAPNSPVVP-----PTPVKNPVKKW-KPPWEDDEPTEEVKKPGE--PEKTPVLA 918
QY 556 KGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPT-----TPKKAP-----KEL 603
DB 919 K-KEPEKPD-APKVAKPRDSSPKKAVPE--KEPAKAAKPRDLSPKALPIPANTQEA 974
QY 604 APTTPGTPS-----TSDKP-----APTTPKET-----APTTPKEPAPTTPKPPAPT 646
DB 975 PPTPVKNPVKWKMPWEDDEPAPVSAPEPEKKTIVLAKKAPAPKPRD--SPKAAV 1031
QY 647 TPEPTPTTSEVSTPTTPKEPTTIHNSP---DESTPELSA-EP---TPKALENSPKREG 698
DB 1032 AAK-PDPKPIEV-PTPVKNPVKWKMPWEDDEPSEVSAPEPEKKTIVLAKKAPKPA 1089
QY 699 V-----PTTKTPPAT-----KPEMTAKKTDTERDRTPEPTT--TAAPK 737
DB 1090 TKPDSEAADPVSGPSSKDKLSKKAPEKPTTPPKDKDLKPSAKKPEKAPAPK 1149
QY 738 MKRETATTTTEKTESKTTATTTQVSTTTQDTTPFKITTLKTTTLAKVTTTKITITTE 797
DB 1150 KKKPVWDDDDDEADTVAPSKKPTDEDPAPLG-----GPKTKDKP----- 1193
QY 798 IMNKPEETAPKDRATNSKATTPPKPOKPTKAPKPKPTSTKKPTMPVRKPK----- 848
DB 1194 -LNKKAPEKPTK-----PKREVSKPEKPTPEPKP-AAKWKMPWEDDEPDE 1243
QY 849 ---TTPPRKMTSTMP 861
DB 1244 ADFTPAPKPKDPTEDP 1259

RESULT 10
09LIER PRELIMINARY: PRT: 1480 AA.
AC 09LIER;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;

Qy	352	EPSPITTKKEAPITTKKEAPITTKKKAAPTITTKKEAPITTKKEAPITTKKKAAPTAKKEAP	411
Db	769	PPALSSPPAPAPQVKKSSPPVQVSSP-PPAKSSPPALAVSSPPQVKTSPPPALSSSP	827
Qy	412	TTPKETAP-----TTP-----KLTPTTPEKLAPTTPEKAP-----TPEELAPITPEEPT	458
Db	828	LAKSSPPHHVSSPPPVVKKSSPPAPVSSPPLTKKRAPAPAHVSSPPEVVKSP-PA	885
Qy	459	PTPEEPAPITTPKAAANTKE-PAITTKKEAPITTKKEAPITTKKEAPITTKKCAIT	517
Db	886	PTTVISP-PSPEKSSPPTPLVSLPPIVKKSSPPAMVSSP-PMTEKSSPPVWVSSPPT	943
Qy	518	LKE---PA-----PTTKK---PAKELAPITTK-----PTTTSOKRAPITTKGAPITP	563
Db	944	VKSSPPAPVSSPPATKSSPPAPVNLPEPVKSSPPTPVSSPPA---PKSSPPAP	1000
Qy	564	KEBAPITPKKEAPITTKGAPITTKKEAPATTKKAPAKELAPITTKGTTSTTSKAPIT	622
Db	1001	MSSPPPEVVKSSPPAPVSSPPPVVKKSSPPAPVSSP---PPVKSPPPPAVSSPPPV	1057
Qy	623	TPKETAPITPKKEAPITTKKAPAPITPEEPPTTSSVSTPTTKKEPTIHKSPPDESTPEL	681
Db	1058	KSPPPAPISPPPVVKKSSPPAPVSS-SPPEVVKSSPPAPVSSPPPIKSPPPAPVSS	1115
Qy	682	SAEPTPALENSPEKGVPTTKPAATKPEMTTAAKDTTEDLRTTPEPT	731
Db	1116	SPPPAP-----VKPEPLP-PPAPVSSPPPVVTPAPPKKEOSSLPPAPSS	1158

RESULT	12		
P70670		PRELIMINARY;	PRT; 2187 AA.
ID	P70670		
AC	P70670;		
DT	01-FEB-1997 (TrEMBLrel. 02, Created)		
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 12, Last annotation update)		
DE	MUSCENP POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC		
DE	MUSCLE-SPECIFIC FORM GP220).		
GN	NACA.		
OS	Mus musculus (Mouse).		
OC	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96312450; PubMed=8698236;		
RA	Yotov W.V., St-Arnaud R.;		
RT	"Differential splicing-in of a proline-rich exon converts alphanMC		
RT	into a muscle-specific transcription factor.";		
RL	Genes Dev. 10:1763-1772(1996).		
DR	EMBL; U48364; AAB18734.1; -;		
DR	EMBL; U48363; AAB18733.1; -;		
DR	MGI; MGI:106095; Naca.		
DR	InterPro: IPR002715; NAC.		
DR	InterPro: IPR003037; TS-N.		
DR	Pfam; PFO1849; NAC; 1.		
DR	Pfam; PFO2094; TS-N; 1.		
SO	SEQUENCE 2187 AA; 22059 MW; 003646AAB64DEBFD CRC64;		

Query Match	15.5%;	Score 776;	DB 11;	Length 2187;
Best Local Similarity	27.7%;	Pred. No. 1.9e-41;		
Matches 288;	Conservative 122;	Mismatches 389;	Indels 240;	Gaps 49;

```

0Y 34 VTPEPTSTTHQHKVSTSKRTTAPINRPSLPENSDPKR-----TSLVNHETT 84
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 806 VQSPKVDPIMSDVTTSTSRKTSATNAV -PK-----DTSATLSLKSVPATYLSLSPKNAV 857
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 85 VETKETTTTTNKQTSJDGEKKTTSAAETOSIEKTSKADLAPTSKYLAFTPKAETTTGPA 144
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 858 AFSNENATVPTETPSLNNALAAAPKETLATSIPKATSPS-----PKPTKSVSLKCAPA 913
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 145 LTT-----PKE-----PTPTTPEKPASTT-----PKPEPTT 171
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db	914	MTSKKATETIAASNDVBSOQPKKVVLLQHVPTSPPKSVSDTLISGALTSPBPKGP-PAT	972
QY	172	IKSAPPTPKK-PAPPTTKKSAPPTK-----EPAPTTKKEAPPTPKKEAPPTTKK	221
Db	973	LAEPPTPKSKSPKPAASKKTPATPSPBEGVAVPLEIPPCSKKAPPTAAKSSAITSSSKR	1032
QY	222	APPT--IKKSAPT-----TKKEAPPTPKK-PAPPTPKKEAPPTPK	257
Db	1033	APKPAVSKLEPSGVTAVPLEISLPKETSNSKATSEKSSAPSKSPKTAGK-----TP	1089
QY	258	KEPPTPKKEAPPTKEAPPTTKKEAPPAKPKAPPT--PKEPAP-----TPKKEAPPT	311
Db	1090	GGVAVAPPELSLPKKEIPQKATPNESLAASQKRSKPTSVKETPPBGGVAVAMPLEIPAP	1149
QY	312	TKESPPTPK-----EPAPTTKKSAPTTKKEAPPTT-KSAPTTK-----EPSP	357
Db	1150	QKAPKTAVPKQIPTBEDAVIILAGSPLSKKASKTAPAKEAPATPSGVIAVSGEISPS	1209
QY	358	TKKEAPPTPKKEAPPTPKKAPPT--PK-----PAPPTPKEPAP-----T	396
Db	1210	KKTSKTAAPKENSITLPPKSPPTAAPKETPATSSGCVAVSELSPPPTPASPAGVVT	1269
QY	397	TTKKAPATKKEAPPTPKKTAATPTPKKLTPTTKEAPPTPK-PAPPTPELAPPTPE	455
Db	1270	LTPKAGANALAE-SPASPKVPKPTAPAEHSTTP-----SPQIPKVAPEKASATPPS	1322
QY	456	EPPTT-----PEEPATTTKKAAPNTPKKEAPPTTKE-PATTTKKEAPPTTKKTA	506
Db	1323	KKTPKTAVPKETSAPSGVAVAPLEIPSPBRAKTAAPKEHPAS--PEGATVAVOIP	1380
QY	507	PTTKGAPPTLKKEAPPTT-----KKAPKELAPTTTKEPT--S	544
Db	1381	PSPKSGSKKAGSKK-PTTTPSPBGTAVAPLEIPISKKTSMASPKKTLVPPSSKLSQT	1439
QY	545	---TSDKAPAPPTPKGTAPPTPKKEAPPTPKK-PAPTTKGTAPTTLKEAPPTPKKPA	599
Db	1440	VGPKEITLEGATVPLEIPPSHKAKRTVDKOVLPTEPK-DAPTTLAE-SPSPSK-A	1496
QY	600	PKELAPTTKGPSTSDKAPAPPTPKETAPPTPKKEAPPTPKKAPPTPEP-----PT	654
Db	1497	PKTAPAPSER-VTTVPPEKPA-TPOKASGTASKVAVPAETOEAVVSSHETPVPAVPV	1554
QY	655	TSEVSTPTT-----EPTTIHKSDESTPELSAPTKALENSPEK-PGVPTTKTPAA	707
Db	1555	KNPSHKKTSTKTELKAPATLPSPSTKSPKIPSSKKAART-SAPKEPAPSPK-PVT	1611
QY	708	TKPEMTTAKDKTTERDLRTPPETTTTAAPKMKETAATTEKTESKITATTTQVYSTTQ	767
Db	1612	T-SLAOTAPPSLOKAPSTTIPKENLAAPAV-----LPVSSKSPAPAPARSASISPATAP	1665
QY	768	DTTPEFKITLITLTKTLTAPKVTTKTKTITTEINMKPEETKPKDRATNSKATPKQKPTK	827
Db	1666	QTAKEKATTTIPSCKKAATETPLEISTAVSLBAGAKEISE-----TSVKVLMSSP----	1716
QY	828	APKPTSTKKTTPKTPVAKKPTPTPRKMTSTPDELNPTSRIAEAMLQTTTPRNPOTNSK	887
Db	1717	-PKAASSKRASTLP-----ATLPSLKEASVLS-----PTATSGK	1752
QY	888	LVEVNPKSEDAAGABGETP	906
Db	1753	DSHISPV-SDACSTGTTP	1770

RESULT	13	
Q9VEL9		
Q9VEL9	PRELIMINARY;	PRT: 2112 AA.
AC	Q9VEL9:	
D1	01-MAY-2000	(TREMBLrel. 13, Created)
D1	01-MAY-2000	(TREMBLrel. 13, Last sequence update)
D1	01-JUN-2001	(TREMBLrel. 17, Last annotation update)
DE	CG4090	PROTEIN.
GN	CG4090	

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyridae; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Gunion R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe C.R., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Adair J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iidegram C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Meriklov G., Mlshina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskiy R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003716; AAF53402.1; -;
 DR FlyBase: FBgn0038492; CG4090.
 DR InterPro: IPR002557; Chitin_binding.
 DR InterPro: IPR000561; EGF-like.
 DR Pfam: PF01607; Chitin_bind_2; 10.
 DR SMART: SM00464; ChitinBD; 11.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS00022; EGF_1; 1.
 SQ SEQUENCE 2112 AA; 219547 MW; B91018E551A5D36 CRC64;

Query Match 14.7% Score 739; DB 5; Length 2112;
 Best Local Similarity 27.2% Pred. No. 4.2e-39;
 Matches 281; Conservative 98; Mismatches 339; Indels 316; Gaps 44;

QY 4 NKKRRTKKRPVVDGAGSGLDNGDFVYTPDTSTGHNKVSPTTKITAKPKNRP 63
 DB 1023 NHPDQIQTKLKKKKVSGGSSN-----STNSSSSSSSNSGSSSS----- 1070
 QY 64 SLPNSTSTKETSITVKKETTV---ETKETTITKKQTS---IDGKETSIAKQTSI-- 114
 DB 1071 SSS 1130
 QY 115 -----EKSSANDLAPTSKVLAKPPTKAEITTKGP----- 143
 DB 1131 NNNNGSS 1190

QY 144 -----ALTTPKKEPTPTTPKEPASTTPKEPTTKISAPTT 178
 DB 1191 SCGGTVMQAOACNHNAMAVKCGSLAPPTSTPTTSR-PTTASISRSDDQSTSRPG 1249
 QY 179 PKEPATTTKSAPTTPKEPAPTTTKKEPAPTTKEPAPTT----- 217
 DB 1250 P-----PTTARPVYARPTTSSTPSSSQTTSPTVQAPNTDGKRSBGFMADPNKSKFYR 1305
 QY 218 -----TKEP-----APT-----TTKSAPTTPKEPAPTTPKKAPPT 248
 DB 1306 CVRNKNGFTSIPGCGAGVWDQIDQTCNHNFNKNCSTGESTTPKAPC--EPATNGTTA 1363
 QY 249 PKEPATTPKEPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAPPAKPAPTTPKEPA 308
 DB 1364 TSSSTTTT--PPTTDLPTSTGTP-PPTTLELPT-----TTTDLPTTTTLRLP 1413
 QY 309 PTTTKEPSPTTPKEPAPTTTKSAPTTTKKEPAPTTTSAPPTPK-----EPSTTTKEPA-- 363
 DB 1414 TTTTSLPPTTTGLPPTTTTGAPPTTTLSESTSTVTTSPESTTQPPSTTMMKPLPAG 1473
 QY 364 -----TTPK-----EPAP---TP 373
 DB 1474 TECTGEGYADPDCRKYTCINAGASYKRYNTCCRGKGMNEVOTCYVENIPKSKL 1533
 QY 374 PKKPAPTP---KEPAPTTPKE-PAPTTTPKPAAPKAPKAPPAPTTPKPKLTP 428
 DB 1534 PAEPITTTPEESKDPGSTPOSTDEPTVTXRI-TPKE-EPSTEKPOKPTQYEPKPT 1591
 QY 429 TPBKLAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAAAPNTKEPAPTTPKE 468
 DB 1592 TTBE--PEKOKPT-TPTEYPOKPTTTEEPKEPOKPT-TPTEYPOKPTTTEEPKEPOK 1647
 QY 489 PAPPT-PEKPAPTTPKETA-----PTTPK--GTAPTTPKEPAPTTPKPAKELAPTTK 540
 DB 1648 PTTTEYPOKPTTEPTTTSIPGNTPTTTSVPGNPTTTPTPVETTTSTPBG---PTTGG 1704
 QY 541 EPTSTSDKPAPTTPKGTATTPKEPAPTTPKE-PAPTTPKGAAPTTPKKEPAPTTPK-- 597
 DB 1705 EPTTTT-LPSTTTDAIQEPTTSKKEPTTTTSSPESSPEGSVTTLQPPQPNVNCSE 1763
 QY 598 -----PAKELA-----PTTKGPTSTSDKPAPTTPKETAAPTTPKKEPAPTTPK 648
 DB 1764 GFPPDPDCSRYYRCVDAANNGYQYAFACGQJWDSTECNADQVSGN----- 1816
 QY 649 ETPPTTSEVSTPTTPKEPTTIKSPDESPPEKSAEPTPALENSPEKPVPTTKPAAT 708
 DB 1817 -----CSSGQTTTGTTEPTTSTSSGKP-----ETTSKAPN-----TTTMA-- 1857
 QY 709 KPEPTTAKDKTTERDLRTPETTTAAPKMTKETATTEKTESKITATTTQVTSSTTOD 768
 DB 1858 -PPTTT-----SSPETTTT--VASETTTTSGT-----TTTATPEPTTKP 1895
 QY 769 TTPPKITTLTKTTLAPKVTITTKTITTTTEIMNKPDEPAKPKKDATNSKATTPKOKTGA 828
 DB 1896 PKP-----ETTTIAGESTTSKSPPTTE-----SPAPSTWTA 1928
 QY 829 PKKPTSTKPKTMP 842
 DB 1929 P-----CPETGP 1935
 RESULT 14
 ID 09XDH2 PRELIMINARY; PRT; 763 AA.
 AC 09XDH2;
 DT 01-NOV-1999 (TEMBLrel. 12, Created)
 DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE PROLINE-RICH MUCIN HOMOLOG.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;

[1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-H37RV:
 RA Espitia C., Lacleite J.L., Mondragon-Palomino M., Martens A.,
 RA Zhang Y., Moreno C., Singh M.:
 RT "Cloning and characterization of a new member of the PGIS family that
 RT is a useful marker of polymorphism in Mycobacterium tuberculosis."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF071081; AAD41594.1; -
 DR InterPro: IPR002951; Atrophin.
 DR InterPro: IPR002965; P_rich_extensin.
 DR InterPro: IPR003882; Pistil_extensin.
 DR PRINTS: PRO1222; ATROPHIN.
 DR PRINTS: PRO1217; PRICEXTENSIN.
 DR PRINTS: PRO1218; PSTLEXTENSIN.
 SQ SEQUENCE 763 AA; 75034 MW; 39168EC45A5916F8 CRC64;

Query Match 14.3%; Score 715; DB 2; Length 763;
 Best Local Similarity 28.2%; Pred. No. 5.8e-38;
 Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

QY 132 PTPEAKETTTGPAITTTKEPPTTTPKEPASTTPEKEPTTTTTSAPTTPKEPAPTTPK 191
 DB 3 PVP-----APRAIAPLPAPAPAEKSKPPPPAP-----PAPCMMLVSAAP 46
 QY 192 TTPEKAPPTTTPKEPAPTTPKEPAPTTPKSAPTTPKKEPAPTTPKKEPAPTTPKE 251
 DB 47 PCP--PAP-----PAPKPKSKAPFPVPVPAPAPAREIAPLP--PAP-----PEAPRE 90
 QY 252 PAPTTPKEPPTT-----PKKEPAPTTPKEPAPTTPK--EPAPTAPK--KPAPTTPKEPAPTTPK 305
 DB 91 SRPALPCPPPVVPIPDPEEAPAPVPAPNSPFPFPFPAPKVPAPVP--FVPSNP 148
 QY 306 EP--APTTPKEPPTTPKEPAPTTPKSAPTTPKAPPTTPKSAPTTPKESPTTPKEPAP 363
 DB 149 FPPPPAPALNPAP-----PAPPLANSPLPAPAPTPAGT--PPAAPVPVAPAPAKSKPA 201
 QY 364 TTPEKAPPTTPKAPPTTPKEPAPTTPKEPAP-----TTTKKAPATAKEPAPTTPKETAP 419
 DB 202 SPPEPPAP-----PMAPTMEFPPLPVPPDPSKETPPAPAPAPIPAPVPPLP 256
 QY 420 TTPEKLTPTTPEKIA-----TTPEKAPPTPEELAPTTPEEPPTTPEEPAPT-- 469
 DB 257 PVPKIKIPAP--APVVAALVAVACPPPLPPLNNHPPAPAPAPVGPVLALEPNSHP 313
 QY 470 ----PKAAPPTTPKEPAP----- 483
 DB 314 PAPPASAVPGVPLAPLPISGRVSVWMSFTTLSTFCRCVSGEVLALNPSSRSPL 373
 QY 484 --TTPEKAPPTTPKEPAPTTPKEPAP-----PTTPKGA-----PTTKKEPAPTTP--KKAPAKE 533
 DB 374 TTTTPALPAIPLPLPLPLPINTAVPIPLPLPVITALAPLPPLPISPVPAP-- 431
 QY 534 IADTTTKEPTSTSDKAPPTTPKGTAPTTKEPAP-----PTTPKEPAPTTPKGTATTT 587
 DB 432 --PIPPKPMPTTTPPLAAPAPBEK--TVVLPDPGSCPPSEKPNPAPPEPEPEKSSAPLP 488
 QY 588 KEAPPTP--KKAPAKELAPTTTGP--TSTSDKAPAPTTPKET--APTTPKEPAPTTP 640
 DB 489 APPAPMSASVAVPSPPIPPAPAPARASMPALPAPAPSPATRLCPPLPSPSPAPNSP 548
 QY 641 KKPAPPTTPPTTSEVSTPTTKEPTTIHKSP-----DSTPELSAEPTPKA--LENSP 694
 DB 549 --PAPAPPTTPPKLLS--ANPPCPVPADNRPAPAPAPAPPLPAPDPPIPPVANS 604
 QY 695 KEKGVPTTKPAATKPEMTTAKDTERDLRTPETTTAPKTKETATTEKTESKI 754
 DB 605 --PAPAPAPAPAPALPVPVNP-----PPTPAPK-----SPAL 637
 QY 755 TAITTOVSTTODTTPFKITTKTTTLADKVTTKKTITTEIMNKPEETAPKORATN 814
 DB 638 PAAPAPAPAPVPVATTP-----PPAPAPAPAPN 665

QY 815 SKATTPKQKP-----TKAPKKTSTKKPKIMPRVKKXTTTPPKKMTSTM-----PELNP 865
 DB 666 SMALPPAPDPPIPLLATTPAPAPPLPMPSPAPPLPAPADPPAPPLTINQPSPLAP 725
 QY 866 TSRIAEMLAQTTPRPNQTPNSKLTVEVNPDKSEADGACGCT 905
 DB 726 VPGAPLAPLPINGRPVARKKSLI-----GSSGDT 756

RESULT 15
 ID O9EP06 PRELIMINARY; PRT; 555 AA.
 AC O9EP06;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE VEGETATIVE CELL WALL PROTEIN GPl.
 GN GPl.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OC NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ferris J.J., Moessner J.P., Waffenschmidt S., Kilz S., Drees J.,
 RA Goodenough U.W.:
 RT "Glycosylated polyproline II rods-with-kinks as a structural motif in
 RT plant hydroxyproline-rich glycoproteins."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF309494; AAC45420.1; -
 DR InterPro: IPR003882; Pistil_extensin.
 DR InterPro: IPR002965; P_rich_extensin.
 DR PRINTS: PRO1217; PRICEXTENSIN.
 DR PRINTS: PRO1218; PSTLEXTENSIN.
 SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;

Query Match 14.1%; Score 707.5; DB 10; Length 555;
 Best Local Similarity 28.9%; Pred. No. 1.3e-37;
 Matches 189; Conservative 66; Mismatches 243; Indels 157; Gaps 22;

QY 194 PKKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKKEPAPTTPKEPAP 253
 DB 40 PPSAPPS--PAPSPAPSP-----APSPAPSPGPPSPAPSPSPS 80
 QY 254 PTPKEPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAPKAPAPTTPKEPAPTTPKEPAPTTPK 313
 DB 81 P-APPSPAPSPAPSPAPSPAPSPAPSPAPSP--APSPAPSPSPAPSPAPSP----- 134
 QY 314 EPSTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPSPPTTPKEPAPTTPKEPAPTTP 373
 DB 135 -PSPSPP-----SPAPPLPSPAPSPS 154
 QY 374 PKKEPAPTTPKEPAPTTPKEPAPTTPKAPATAKEPAPTTPKETAPTTPKEKLTPTPEKL 433
 DB 155 PSPVPVPSPPVPVPPSPAPSPSP-----PSSSPVPVPPSPAPSP----- 194
 QY 434 APTTPKEPAPTTPKEELAPTTPEEPPTTPEEPAPTTPKAAAPTTPKEPAPTTPKEPAPTTP 493
 DB 195 APPVPSAPAPSP--APVPSAPAPSPSPAPSPSPAPSPSPAPSPSPAPSPVPSP- 250
 QY 494 PKKEPAPTTPKEPAPTTPKGTAPTTKEPAPTTPKKAAPKLAATTTKEPTSTSDKAPAP 553
 DB 251 PPSAPPSAPPPAP-----PPPSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 298
 QY 554 TPKGTAPTTTPKEPAPTTPKEPAPTTPKGTAPTTKEPAPTTPKKAAPKELAPTTTGPS 613
 DB 299 PPTPPPPSPSPSP--VPSPAPVPSPAPSP-----PAPSPSPAP-----PSPSPSP 349
 QY 614 TTSKAPAPTTPKKEPAPTTPKEPAPTTPKAPAPTTPETPTTSEVSTPTTKEPTTIHKS 673
 DB 350 SPSPSPSPSP--PSPSPSPSPSP--PSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 393


```

QY 674 PDESTPELSAEP-PALENSPEKPGVPTTKTPAATKPEMTTAKDKTERDLRTTPEY 732
Db 394 ---AFDDLNGTSTRPGSASRMVGEPIAGTKCKGNLKGMPKPSRNPBGQAVFSGRTV 450
QY 733 TAAPKMTKETATTEKTESKITATTOVTSTTODTPEKITTLKTTLAPKVTTK-- 790
Db 451 GSVANTIRVAFATEK--PALIYSSIELVYNTG-----ATLIRVIAANVTRSQIR 500
QY 791 ---KITTTIIMNKPEETAKPKDRATNSKAT-PPQKPTKAPKKPTSTKKPKT 840
Db 501 CPGLTYGCTTPIAGVP-----TGIDATTWPNWKIAGVRINMGAGNKKPKT 545

```

Search completed: April 26, 2002, 16:21:42
 Job time: 487 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:28:53 ; Search time 210.7 Seconds

(without alignments)
946.223 Million cell updates/sec

Title: AA2
Perfect score: 7276
Sequence: 1 MANKTLPYLLLLSVFVIO.....ARATRRSGQLSKWVYVNCNP 1363

Scoring table: BL0SUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP_ARCHAEA:17:1
2: SP_BACTERIA:17:2
3: SP_FUNGI:17:3
4: SP_HUMAN:17:4
5: SP_INVERTEBRATE:17:5
6: SP_MAMMAL:17:6
7: SP_MHC:17:7
8: SP_ORGANELLE:17:8
9: SP_PHASE:17:9
10: SP_PLANT:17:10
11: SP_PODONT:17:11
12: SP_VIRUS:17:12
13: SP_VIRTEBRATE:17:13
14: SP_UNCLASSIFIED:17:14

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7245.5	99.6	1404	4 Q92954	Q92954 homo sapien
2	7216.5	99.2	1404	4 Q9BX49	Q9BX49 homo sapien
3	7116.5	51.1	1054	11 Q9JMH9	Q9JMH9 mus musculu
4	1713.5	23.6	401	6 077765	077765 bos taurus
5	982.5	13.5	1079	5 Q9N4S7	Q9N4S7 caenorhabd
6	946.5	13.0	1049	5 Q917S1	Q917S1 drosophila
7	892	12.3	1795	5 076894	076894 drosophila
8	863.5	11.9	1489	10 Q96449	Q96449 phytothor
9	846	11.6	1274	5 Q20007	Q20007 caenorhabd
10	817	11.2	1480	10 Q9LIR8	Q9LIR8 arabidopsis
11	816	11.2	1315	10 Q9SPM6	Q9SPM6 zea mays (m
12	812	11.2	2187	11 P70670	P70670 mus musculu
13	786.5	10.8	2112	5 Q9VEL9	Q9VEL9 drosophila
14	784	10.8	1188	10 Q41805	Q41805 zea mays (m
15	775	9.8	763	2 Q9YXH2	Q9YXH2 mycobacteri
16	709.5	9.8	6677	5 Q9N435	Q9N435 caenorhabd
17	707.5	9.7	555	10 Q9FPQ6	Q9FPQ6 chlamydomon
18	703.5	9.7	4880	11 Q9JLTI	Q9JLTI rattus norv
19	703.5	9.7	5085	11 Q9JMS6	Q9JMS6 rattus norv

20	695	9.6	1151	13 Q57580	Q57580 gallus gall
21	684.5	9.4	1229	5 Q94185	Q94185 caenorhabd
22	674.5	9.3	2089	4 Q14676	Q14676 homo sapien
23	667	9.2	2284	5 Q9VPG1	Q9VPG1 drosophila
24	666.5	9.2	7962	4 Q10465	Q10465 homo sapien
25	665	9.1	4833	11 Q9YXK6	Q9YXK6 mus musculu
26	665	9.1	5038	11 Q9YXK7	Q9YXK7 mus musculu
27	661.5	9.1	2768	5 Q9VC00	Q9VC00 drosophila
28	653.5	9.0	3570	4 Q99552	Q99552 homo sapien
29	653	9.0	3570	4 Q99552	Q99552 homo sapien
30	643.5	8.8	990	13 Q91803	Q91803 xenopus lae
31	636	8.7	3507	5 Q23587	Q23587 caenorhabd
32	636	8.7	6632	5 Q17362	Q17362 caenorhabd
33	633	8.7	489	10 Q41707	Q41707 vigna ungu
34	632	8.7	761	10 Q9Z010	Q9Z010 arabidopsi
35	631	8.7	2344	5 Q9N3Y8	Q9N3Y8 caenorhabd
36	629	8.6	6642	5 Q01761	Q01761 caenorhabd
37	628.5	8.6	1612	5 Q9VVO2	Q9VVO2 drosophila
38	627.5	8.6	971	5 Q9XVS4	Q9XVS4 caenorhabd
39	622.5	8.6	839	2 Q9KX57	Q9KX57 delnoccocus
40	607.5	8.3	801	5 Q23635	Q23635 caenorhabd
41	607.5	8.3	924	12 Q99307	Q99307 epstein-bar
42	605	8.3	379	5 Q27929	Q27929 drosophila
43	600.5	8.3	1893	5 Q9NKC9	Q9NKC9 drosophila
44	598.5	8.2	409	10 Q9SBM1	Q9SBM1 volvox cart
45	591.5	8.1	2232	5 P91365	P91365 caenorhabd

ALIGNMENTS

RESULT 1
ID Q92954 PRELIMINARY: PRT: 1404 AA.
AC Q92954;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MEGAKARYOCYTE STIMULATING FACTOR.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calveti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhalla S., Kriz R., Hewick R., Clark S.C.,
RA Blood 78:279-279(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,
RA Fitzgerald M., Scalitro J., Kelleher K., Preissner K., Kriz R.,
RA Jacobs K., Turner K.,
RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,
RL Mosher D.F. (eds.);
RL Biology of vitronectins and their receptors., pp.45-52,
RL Elsevier Science Publishers B.V. (1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,
RA Kelleher K., Giannotti J., Calveti J., Fitzgerald M., Kriz M.J.,
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,
RA Bhalla S., Kriz R., Hewick R., Clark S.C.,
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
EMBL: U70136; AAB09089.1; -
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR001212; Somatomedin_B.
DR InterPro: IPR002400; GF_cysknot.
DR Pfam: PF00045; hemopexin; 2.
DR Pfam: PF01033; Somatomedin_B; 2.
DR PRINTS: PR00438; GFCYSKNOT.

PRINTS: PRO0022; SOMATOMEDINB.
 DR PROSITE: PS00024; HEMOXEMIN: UNKNOWN_1.
 DR PROSITE: PS00524; SOMATOMEDIN_B: 2.
 DR SMART: SM00120; HK: 2.
 DR SMART: SM00201; SO: 2.
 SO SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Query Match 99.6%; Score 7245.5; DB 4; Length 1404;
 Best Local Similarity 97.1%; Pred. No. 0;
 Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

1 MAMKTLPIYLLLLSVFVIOOVSSO----- 25
 1 MAMKTLPIYLLLLSVFVIOOVSSO----- 25
 26 -----ELSCGRCFESFESFEGRECDCAOCCCKYDKCCPDYIESCAEYHNTSPSSKKAP 79
 61 KRVCTAELSCGRCFESFESFEGRECDCAOCCCKYDKCCPDYIESCAEYHNTSPSSKKAP 120
 80 PPSGASQTIKSTTRSPKPPKPKKKTKKYIESEETIEHSHSVSENOESSSSSSSSSTIW 139
 121 PPSGASQTIKSTTRSPKPPKPKKKTKKYIESEETIEHSHSVSENOESSSSSSSSSTIW 180
 140 KIKSSKSNANRELQKLLKLVKNKKNTKKKPPKPPVVDVAGSGLDNGDFKVTPTST 199
 181 KIKSSKSNANRELQKLLKLVKNKKNTKKKPPKPPVVDVAGSGLDNGDFKVTPTST 240
 200 TQHNKVSSTPKITAKPKNRPSLPPNSDSTKETSIVNKKETVETKETTITNKQSTDG 259
 241 TQHNKVSSTPKITAKPKNRPSLPPNSDSTKETSIVNKKETVETKETTITNKQSTDG 300
 260 KEKTSAKETOSIEKTSADLAPTSKVLAKPPTKAEETTKGAPALTTPKEPTTPPKBAS 319
 301 KEKTSAKETOSIEKTSADLAPTSKVLAKPPTKAEETTKGAPALTTPKEPTTPPKBAS 360
 320 TTPKPEPTTITKSAPTTKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTK 379
 361 TTPKPEPTTITKSAPTTKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTK 420
 380 APTTTSKAPTTKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPK 439
 421 APTTTSKAPTTKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPK 480
 440 BPAPTPAKKPAATTTKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTK 499
 481 BPAPTPAKKPAATTTKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTK 540
 500 TTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPK 559
 541 TTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPK 600
 560 APTAPKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTT 619
 601 APTAPKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTT 660
 620 PEPBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTK 679
 661 PEPBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTK 720
 680 APTTPKPAKELAPTTKEPTSTSDKAPPTPKGATTPKEBPATTTKSAPTTPKEBPATTT 739
 721 APTTPKPAKELAPTTKEPTSTSDKAPPTPKGATTPKEBPATTTKSAPTTPKEBPATTT 780
 740 TAPTTTKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTT 799
 781 TAPTTTKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTTKSAPTTPKEBPATTT 840
 800 KPAPTTPETPTTSEVSTPTTKEPTTTHKSPEDESTPELSAEPPTKALENSPKKEGVPT 859
 841 KPAPTTPETPTTSEVSTPTTKEPTTTHKSPEDESTPELSAEPPTKALENSPKKEGVPT 900
 860 TKTPPAATKPEMTTAKDKTTERDLRTTPEPTTAAPKMKETATTTTEKTESKITATTTOV 919

901 TKTPPAATKPEMTTAKDKTTERDLRTTPEPTTAAPKMKETATTTTEKTESKITATTTOV 960
 920 TSTTQDTTPPKKITTLTKTTTLAPKVTYTTKKTITTEIMNKEEELAKPKDRATNSKATTPK 979
 961 TSTTQDTTPPKKITTLTKTTTLAPKVTYTTKKTITTEIMNKEEELAKPKDRATNSKATTPK 1020
 980 POKPTKARKKPTSTKKPTMPVRKPKTTPTRKMTSTMPELNPTSRJAEAMLOTTTRPN 1039
 1021 POKPTKARKKPTSTKKPTMPVRKPKTTPTRKMTSTMPELNPTSRJAEAMLOTTTRPN 1080
 1040 QPNSKLVENVNPKSDAGAEGETPHMLLRPHVMEVTPPMDYLPRVNOGIINPMLS 1099
 1081 QPNSKLVENVNPKSDAGAEGETPHMLLRPHVMEVTPPMDYLPRVNOGIINPMLS 1140
 1100 DETNICNGKPVLDGLTTLNGLTIVARGHYFWMLSFSPSPSPARRRTTEWGLPSPTDYFT 1159
 1141 DETNICNGKPVLDGLTTLNGLTIVARGHYFWMLSFSPSPSPARRRTTEWGLPSPTDYFT 1200
 1160 RCNCEKTEFEFDOSQWFRFTNDIKDAGYKPIFKGFGGLTGOIYVAAALSTAKYKMMESVY 1219
 1201 RCNCEKTEFEFDOSQWFRFTNDIKDAGYKPIFKGFGGLTGOIYVAAALSTAKYKMMESVY 1260
 1220 FFKRGSSIQOYIYKQEPVOKCGRBPALNYPYGEWQYRRRRFRRAIGPSQTHTRIQY 1279
 1261 FFKRGSSIQOYIYKQEPVOKCGRBPALNYPYGEWQYRRRRFRRAIGPSQTHTRIQY 1320
 1280 SPARLAYQDKGVJHNVKXSIIMRGILPNVTSALSPNTRKPDGYDYAFSDQYNNIDV 1339
 1321 SPARLAYQDKGVJHNVKXSIIMRGILPNVTSALSPNTRKPDGYDYAFSDQYNNIDV 1380
 1340 PSRTARAITRRSGQILSKWYNCP 1363
 1381 PSRTARAITRRSGQILSKWYNCP 1404

RESULT 2
 ID 09BX49 PRELIMINARY; PRT: 1404 AA.
 AC 09BX49;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
 DE B6174L6.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR).
 GN B6174L6.2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wray P.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL133553; CAC36090.1;
 SO SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEB5 CRC64;

Query Match 99.28%; Score 7216.5; DB 4; Length 1404;
 Best Local Similarity 96.8%; Pred. No. 0;
 Matches 1359; Conservative 0; Mismatches 4; Indels 41; Gaps 1;

1 MAMKTLPIYLLLLSVFVIOOVSSO----- 25
 1 MAMKTLPIYLLLLSVFVIOOVSSO----- 25
 26 -----ELSCGRCFESFESFEGRECDCAOCCCKYDKCCPDYIESCAEYHNTSPSSKKAP 79
 61 KRVCTAELSCGRCFESFESFEGRECDCAOCCCKYDKCCPDYIESCAEYHNTSPSSKKAP 120
 80 PPSGASQTIKSTTRSPKPPKPKKKTKKYIESEETIEHSHSVSENOESSSSSSSSSTIW 139
 121 PPSGASQTIKSTTRSPKPPKPKKKTKKYIESEETIEHSHSVSENOESSSSSSSSSTIW 180

```

QY 140 KIKSSKNSAANRELQKLVKDNKKNRKTKKPPKPVVDEAGSLDNGCFVTTPDST 199
DB 181 KIKSSKNSAANRELQKLVKDNKKNRKTKKPPKPVVDEAGSLDNGCFVTTPDST 240
QY 200 TOHNKVSYPKTTTAKPVPNSLSPNSDTSKETSIVNKEETVETKETTNNKOTSTDG 259
DB 241 TOHNKVSYPKTTTAKPVPNSLSPNSDTSKETSIVNKEETVETKETTNNKOTSTDG 300
QY 260 KEKTSIAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTGKALTPKEPTTPKEAS 319
DB 301 KEKTSIAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTGKALTPKEPTTPKEAS 360
QY 320 TTPKEPTPTTTSAPTTKEPAATTTKSAPTTPKEPAATTTKEPAATTTKEPAATTTKEP 379
DB 361 TTPKEPTPTTTSAPTTKEPAATTTKSAPTTPKEPAATTTKEPAATTTKEPAATTTKEP 420
QY 380 APPTTKSAPTTKEPAATTTKSAPTTPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEP 439
DB 421 APPTTKSAPTTKEPAATTTKSAPTTPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEP 480
QY 440 EPAPTPAKKPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPA 499
DB 481 EPAPTPAKKPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPA 540
QY 500 TTKSAPTTKEPSPTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPA 559
DB 541 TTKSAPTTKEPSPTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPA 600
QY 560 APAPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEP 619
DB 601 APAPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEP 660
QY 620 PEEPAATTTKAAADNTPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPA 679
DB 661 PEEPAATTTKAAADNTPKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTTKEPA 720
QY 680 APPTPKKPAPELAPTTTKEPTSTSDKPAATTTKGAATTTKPAATTTKEPAATTTKEPAATTTKEP 739
DB 721 APPTPKKPAPELAPTTTKEPTSTSDKPAATTTKGAATTTKPAATTTKEPAATTTKEPAATTTKEP 780
QY 740 TAPPTLKEPAATTTKPAPELAPTTTKEPTSTSDKPAATTTKGAATTTKPAATTTKEPAATTTKEP 799
DB 781 TAPPTLKEPAATTTKPAPELAPTTTKEPTSTSDKPAATTTKGAATTTKPAATTTKEPAATTTKEP 840
QY 800 KPAPTTKEPTPTTSEVSTPTTKEPTTTIKSDSESTPELSAETPKALENSKEPGVPT 859
DB 841 KPAPTTKEPTPTTSEVSTPTTKEPTTTIKSDSESTPELSAETPKALENSKEPGVPT 900
QY 860 TKTPAATKPEMTTAKDKTERDLRTPTETTTAARKKETAATTTTEKTESKITATTTQV 919
DB 901 TKTPAATKPEMTTAKDKTERDLRTPTETTTAARKKETAATTTTEKTESKITATTTQV 960
QY 920 TSTTTQOTTEPKTTTLTKTTLAKRVTTTKTTTTEIMNKPEETAKPKDRATNSKATTPK 979
DB 961 TSTTTQOTTEPKTTTLTKTTLAKRVTTTKTTTTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 980 POKPTAPKPKPTSTKPKTPVARKPKTPTTPKMTSTMBELNPTSRIEAMQOTTRN 1039
DB 1021 POKPTAPKPKPTSTKPKTPVARKPKTPTTPKMTSTMBELNPTSRIEAMQOTTRN 1080
QY 1040 QTNNSKLVENPKSEDAAGAGETPHMLLRPHVEMPEVTPDMQDYLPRVNOGIIINPMIS 1099
DB 1081 QTNNSKLVENPKSEDAAGAGETPHMLLRPHVEMPEVTPDMQDYLPRVNOGIIINPMIS 1140
QY 1100 DETNINCKRPVQDLTLTRNGTLVAFRGHYFWMISPPSPSPARITTEVIGISPIDTYVT 1159
DB 1141 DETNINCKRPVQDLTLTRNGTLVAFRGHYFWMISPPSPSPARITTEVIGISPIDTYVT 1200
QY 1160 RCNCEKTEFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIVALASTAKKKNPESEY 1219
DB 1201 RCNCEKTEFFFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIVALASTAKKKNPESEY 1260
QY 1220 FFKRGSIQOYIYKQEPVQKCGRRPALNVPYGEHTOVRRRRFRERAIIGSOTHTIRIQY 1279

```

```

DB 1261 FFKRGSIQOYIYKQEPVQKCGRRPALNVPYGETTOVRRRRFRERAIIGSOTHTIRIQY 1320
QY 1280 SPARLAYODKVLHNEKVSITLMRGLPNVYVTSALSPINIKKPDGYDYAFSKOYINIDV 1339
DB 1321 SPARLAYODKVLHNEKVSITLMRGLPNVYVTSALSPINIKKPDGYDYAFSKOYINIDV 1380
QY 1340 PSRTARAITTRSGGTLISKVWYNC 1363
DB 1381 PSRTARAITTRSGGTLISKVWYNC 1404

RESULT 3
Q9JM99 PRELIMINARY; PRT; 1054 AA.
ID Q9JM99
AC Q9JM99;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DI 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DI 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE mRNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN PRG4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikegawa S., Nakamura Y.,
RT "a novel mouse gene highly homologous to a human gene encoding
RT megakaryocyte stimulating factor precursor and cartilage superficial
RT zone protein."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB034730; BAA92310.1; .
DR MGI; MGI:1891344; Prg4.
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR001212; Somatomedin_B.
DR Pfam: PF01033; Somatomedin_B; 2.
DR PRINTS: PR01217; PRICHEXTNSN.
DR PRINTS: PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1054 AA; 115991 MW; 4fC64BFA42283235 CRC64;

Query Match 51.1%; Score 3716.5; DB 11; Length 1054;
Best Local Similarity 54.5%; Pred. No. 2,2e-233;
Matches 772; Conservative 58; Mismatches 171; Indels 415; Gaps 26;

QY 1 MAWKLPYIILLLSVFIQOYSSQ----- 25
DB 1 MGWKLPVCLSLFLPVLIQOVSQDLSSCAGRCGEGYSRDATCNCNDYNCQHYMECCPDF 60
QY 26 -----ELSCGRCESEFERGECDDAQCKKYDCCEYDFCAEVNPTSPSSKKAP 79
DB 61 KRVCSPELSCGRCESEFARRGREDCDSQCKQYKCCADYDFCEEVHNSYS-PSSTAP 119
QY 80 PPSGASQIITKSTTKSKPKPPKPKTKKVIIESEITEHSHSVSENOSSSSSSSSSTIW 139
DB 120 TPAGASDITKSTTKSKPKSPPT-TRTIKVESEELTEHSHSENGE--SSSSSSSSSTIR 176
QY 140 KIKSSKNSAANRELQKLVKDNKKNRKTKKPPKPVVDEAGSLDNGCFVTT--TPDT 197
DB 177 KIKSSKNS-ANRELQKNPNVANDKKNTPKKPNPEPAVDEAGSLDNGCFVTTPEPDP 235
QY 198 STOHNKVSYPKTTTAKPVPNSLSPNSDTSKETSIVNKEETVETKETTNNKOTSTDG 257
DB 236 PTPHSHKVSYPKTTTAKPVPNSLSPNSDTSKETSIVNKEETVETKETTNNKOTSTDG 295
QY 258 DGKEKTSIAKETOSIEKTSKADLAPTSKVLAKPTPKAETTTGKALTPKEPTTPKEP 317

```

```

Db 296 -SKKTKTSAKTSKSDKV-----EP 319
OY 318 ASPTTPEPTTITKSAPTTKEBPATTTKSAPTTPKEBPATTTKEBPATTTK 377
Db 320 TSTTPK-----NSAPTTTKRPV-TTKESKPLP-----LPOEPTTAK 357
OY 378 EPAPTTSKAPTTPKEBPATTTKPKAPTTKPKAPTTKPEPTTTPKEBPATTT 437
Db 358 EPPTTKKPEPTTKKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTT 417
OY 438 PKBPATPKAPTTTPKEBPATTTTPKEBPATTTKEBPATTTKEBPATTTKSAPT 497
Db 418 PKEGPTTPKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBPATTTKEBP 461
OY 498 PTTTSKAPTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPK 557
Db 462 -----EPPTPKPEPTTKKEBPATTTKEBPATTTTPKEBPATTTKEBPATTT 512
OY 558 KPAPTAKBPATTPKETAAPTTPKLTTPPEKLTAPTTPKEBPATTTPEELAPTTP 617
Db 513 -----512
OY 618 TPPEBPATTPKKAAPTTPKEBPATTTTPKEBPATTTTPKEBPATTTTPK 677
Db 513 -----TPKEBPATTTTPKEBPATTTTPKEBPATTTTP-----K 539
OY 678 EPAPTTPKPKAPKELAPTTTKEPTSTSDKAPATTPKGTAPTTPKEBPATTTTP 737
Db 540 EPPTTPKAP-----ETTPKEBPATTTTPKEBPATTTTP-----570
OY 738 KGTAPTTPKEBPATTPKPKAPKELAPTTTGTSTSDKAPATTTTPKEBPATTT 797
Db 571 -----KEBPATTPKEBPATTTTPKEBPATTTTPKEBPATTTTPKEBPATTT 609
OY 798 PKKAPTTPPEPTTTPSTSVPTTTPKEPTTIHKSPEDESTPLSAEPKALENSPK 857
Db 610 PKKPEPTT-----617
OY 858 PTTKTPATKPEMTTAKDKTTENDLRTPTTPPTTAAPKMTKETATTEKTESKITATT 917
Db 618 -----TSPKT-----622
OY 918 OVTSTTTODTTPKKTITLTKTTLAPKYTTTKITTEIMNKPPEELAPKPKRATNSKAT 976
Db 623 -----TTTKATTLAPKYTAPAE-----ETQNKPEPTTPASDSODSKITL 662
OY 977 -----TPKP-QKPTKAPKPTSTKPKTTPMPVRKPKTTPTPPKMTSTMBELNPTSRIA 1028
Db 663 KQKPTKAPKPTTKKPTKAPKPTSTKPKPT--PKTRKPKTTPAPLKTTSATPELMTTP--L 719
OY 1029 EAMLOTTPRNOTPNSKLVENPKSEDAAGAGETPHMLLRPHVEMPEVTPDMOYLPRVP 1088
Db 720 EAMLPTTTPKQTPNPEAEVNPDEHDDGGEKRP-LPGRPVLPPTALFETDLAAGRL 778
OY 1089 NGGIINPMLSPETNICKNGKPYDGLTTLNGLTVAFRGHYFWMLSFSPSPARRITEVW 1148
Db 779 NGGININPSPDETNICKNGKPYDGLTTLNGLTVAFRGHYFWMLSFSPSPARRITEVW 838
OY 1149 GISPSPIDVFTTRCNCGEKTFEKKDSQYWFNTNDIDAGYPKPIFGFGLGQIYAALST 1208
Db 839 GISPSPIDVFTTRCNCGEKTFEKKDSQYWFNTNDIDVDPGPKQIVGFGGLGQIYAALST 898
OY 1209 AKYKMPPEVYFFKRGSGTQOYIYKQEPQKCGRRPALNTPVYGEMLTVRRRRFERAIG 1268
Db 899 AKYKDPPEVYFFKRGSGTQOYIYKQEPQKCGRRPALNTPVYGEMLTVRRRRFERAIG 958
OY 1269 PSQTHTRIOYS-PARLAYQDKGVLANEVKYSILMGRLPNVYTSALSLPNIRKPDGY 1327
Db 959 PRQTHTRFHYSVPMKVSVDQKGLFANEVKSITMMGFPNVYTSALSLPNIRKPDGY 1018
OY 1328 AFSKQOYINIDVPSRTARATITTRSGQTLISKWYNCP 1363

```

```

Db 1019 AFSKQOYINIDVPSRTARATITTRSGQTLISKWYNCP 1054
RESULT 4
ID 07765 PRELIMINARY: PRT: 401 AA.
AC 07765;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SUPERFICIAL ZONE PROTEIN (FRAGMENT).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ARTICULAR CARTILAGE;
RA Schumacher B.L., Hughes C.E., Kuettner K.E., Caterson B.,
RA Aydelotte M.B.;
RT *Immunodetection and Partial cDNA Sequence of the Proteoglycan,
RT Superficial Zone Protein, Synthesized by Cells Lining Synovial
RT Joints";
RL J. Orthop. Res. 0:0-(1998).
DR EMBL; AF056218; AAD13404.1; -.
DR InterPro; IPR000583; Hemopexin.
DR Pfam; PF00045; hemopexin; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR SMART; SM00120; HX; 2.
FT NON_TER 1
SQ SEQUENCE 401 AA: 44952 MW: 86147CC9AFB73D7 CRC64:

Query Match 23.6%; Score 1713.5; DB 6; Length 401;
Best Local Similarity 80.2%; Pred. No. 9.3e-104;
Matches 320; Conservative 24; Mismatches 52; Indels 3; Gaps 3;

OY 966 PKRATNSKATTPKPKAPKPTSTKPKTTPMPVRKPKTTPPK-MISTMBELNPT 1024
Db 5 PKRATNSQVTTTPKPKAPKPTSTKPKPT--PVRKPKTTPPKTTSAMEBPPT 63
OY 1025 SRTAEAMLQTTTPRNOTPNSKLVENPKSEDAAGAGETPHMLLRPHVEMPEVTPDMOYL 1084
Db 64 S-LPEAMLQTTTPRPTTPNSELIDVNSENDGAAAGSEPHMIFRPVLTPIVIGETELI 122
OY 1085 PVPNNGIINPMLSPETNICKNGKPYDGLTTLNGLTVAFRGHYFWMLSFSPSPARRI 1144
Db 123 VRGSPQGFGLNPMFSPDETNICKNGRPYDGLTTLNGLTVAFRGHYFWMLTPTTPPPPRRI 182
OY 1145 TEVWGISPIDVFTTRCNCGEKTFEKKDSQYWFNTNDIDAGYPKPIFGFGLGQIYA 1204
Db 183 TEVWGISPIDVFTTRCNCGEKTFEKKDSQYWFNTNDIDAGYPKLISKQFGGLNGKIYA 242
OY 1205 ALSTAQKKNMPEVYFFKRGSGTQOYIYKQEPQKCGRRPALNTPVYGEMLTVRRRRFE 1264
Db 243 ALSTAQKKNMPEVYFFKRGSGTQOYIYKQEPQKCGRRPALNTPVYGEMLTVRRRRFE 302
OY 1265 RAIGPSQTHTRIRIYSPARLAYQDKGVLANEVKYSILMGRLPNVYTSALSLPNIRKPDGY 1324
Db 303 RAIGPSQVHTIRIHYTPVRVYQDKGFLANEVYSTLMRGLPNVYTSALSLPNIRKPDGY 362
OY 1325 DYAFSKQOYINIDVPSRTARATITTRSGQTLISKWYNCP 1363
Db 363 DYAFSKQOYINIDVPSRTARATITTRSGQTLISKWYNCP 401

```

```

RESULT 5
ID 09N4S7 PRELIMINARY: PRT: 1079 AA.
AC 09N4S7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

```

DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE Y51B11A.1 PROTEIN.
 GN Y51B11A.1
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 NC NCBL_Taxid=6239;
 RN 111
 RC SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT Genome sequence of the nematode *C. elegans*: a platform for
 RT investigating biology. The *C. elegans* Sequencing Consortium.;
 RT Science 282:2012-2018(1998).
 RN 121
 RC SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Coulton M.;
 RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN 131
 RC SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN 141
 RC EMBL AC006797; AAE60743.1;
 DR InterPro; IPR002965; P_rich_extensn.
 DR PRINTS; PRO1217; PRICHEXTENS.
 SQ SEQUENCE 1079. AA; 110532 MW; 8BDE3824CF80CA1 CRC64;

Query Match 13.5%; Score 982.5; DB 5; Length 1079;
 Best Local Similarity 29.6%; Pred. No. 7,7e-56;
 Matches 332; Conservative 131; Mismatches 481; Indels 177; Gaps 44;

QY 154 OKLTKVKNKNNKRRKPKPPVVDKSGLDNGDFV-----TTPDSTGQ-HNKV 205
 DB 27 OKLEIACDRCANQPRHTMLP-----STLTVSDMETPSTLVLSSTPSSSTPIKETT 77
 QY 206 STSPKITAARINRPSLRPNSDTSKETSLTVNKEET-VETKETTNNKOSTDGEKETT 264
 DB 78 TTAETTTSTP-----PSSSTTPVQTTTAAETTTSTAPSSSTTPVQTT-----TT 124
 QY 265 SAKETOSIEKTSADKLAPTSKVLAKPTPKAETT-TKGP-ALTPKKEPTTTPKEPASTTP 322
 DB 125 TAPETTSPEPSSS-----TSPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEP 180
 QY 323 KEPTTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 382
 DB 181 PSSSTSPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEP-TTSTEP-PSSSTTPVQTT 236
 QY 383 TTKCAPTPKEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 442
 DB 237 TTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEP-PSSSTTPVQTTTAAETTTSTEP 295
 QY 443 PTAKKKAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKKAAPTTP 488
 DB 296 PSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEPSSSTTPVQTTTAAETTT 355
 QY 489 APTTTK--EAPTTTSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKKAAPTTPKKAAPTTP 545
 DB 356 APETTTKEPSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEP-----APT 411
 QY 546 TTPKEPAPTTPKKAAPTTPKEPAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 603
 DB 412 SSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEP-----PSSSTTPVQTT 467
 QY 604 PEELAPTPPEEPPTTPPEEPAPTTPKAAAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKAAPT 661
 DB 468 TTTAETTTSTEP-PSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEP-TTST 525
 QY 662 ETAPTPKGAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 720

DB 526 PSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTT 565
 QY 721 KPEAPAPTTPKEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 777
 DB 586 APE---TTSTPSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTAPSSSTTP 642
 QY 778 APTTPKEPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 824
 DB 643 SSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEP 702
 QY 825 PTTHKSPDESTPRLSAPF--PKALNSKRECVPTTPKPAATKPEMTTAAKDKTTERO 882
 DB 703 ETTSTPSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEP 762
 QY 883 LRTT---PETTTAARPKMETATTEKTESKITATTTQVSTTTQDTPPFKITTLLKTT 938
 DB 763 VQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEPSSSTTPVQTTTAAETTTSTEP 820
 QY 939 TIAKVTITTKITTTTETIINKPEETA-KPRDRTNSKATTPKPKOKPKPKPTSKPKK 997
 DB 821 STEPPSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEP 878
 QY 998 TMAPVRKPTTPPKMTSMPELNPSTRIAEMLQTTTPRNPQTPSKLVEVNPKSDEAG 1057
 DB 879 QTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEP-PSSSTTPVQTTTAAETTTSTEP 932
 QY 1058 GAGC-ETPHMLRPHVMEVP--DMDYLPV-----PNOGIINPMUSDEN 1103
 DB 933 ETTSTPSSSTTPVQTTTAAETTTSTAPSSSTTPVQTTTAAETTTSTEPSSSTTPVQTTTAAETTTSTEP 992
 QY 1104 -----ICNGKRVDDLTLRNGTLVAFFGHYFWMLS-----PFSPPS 1139
 DB 993 AEFVSTFDIGCTAILTCIYSEGISML-NATL-----FIGLSDGSSIDLFP----- 1040
 QY 1140 PAARITEVWGISPIDTFTRCNCEGKTFEKKDQVWRETN 1180
 DB 1041 -----YVNPGLLE-IMPEINCEBKN-----WSINN 1064

RESULT 6
 ID 0917S1 PRELIMINARY; PRT; 1049 AA.
 AC 0917S1;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE CG5328 PROTEIN.
 GN CG5328.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBL_Taxid=7227;
 RN 111
 RC SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brooksstein P., Brotlier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu L.E., Center A., Chandra I.,
 RA Cherry J.M., Caley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., May A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fierla S., Fleischmann W.,

RA Foster C., Gabriellan A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshirefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spindler A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,
 RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "the genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003495; AAG22353.1;
 RL Flybase: FBgn0030561; CG5228.
 SQ SEQUENCE 1049 AA; 107278 MW; 954DD629E7619671 CRC64;

Query Match 13.0%; Score 946.5; DB 5; Length 1049;
 Best Local Similarity 34.4%; Pred. No. 1.6e-53;
 Matches 376; Conservative 69; Mismatches 480; Indels 169; Gaps 56;

OY 65 EVHNPSTSPSSKKAPPASGASQITKTKRSPKPNKKYKVISEITHEHSSENOE 124
 DB 33 EIGTAKPTTLTKPTGTSAKPTTLKPTGTSAKPTTLKPT-----EGTAKPTTLKPT 86
 OY 125 SSSSSSSSSSTIMKIKSSKNSANRELKLLKDKMKNKTKKPPKPPVDEAGSG 184
 DB 87 GTYAKPTT-----LKPTEGTSK-----KPTTLKPT-----GTAKPTTLKPT---EGTS 128
 OY 185 LNDGDKVTPDPS--TIOHKNVSTSPKITAKPINRPSLPSNDSKETSIVNKEET 242
 DB 129 AKPTTLKPTGTAKTTLKPTGTSAKPTTLKPTGTSAKP---TTLKPTGTSAKPTT 185
 OY 243 VETRETTTNNKOT--STDGKEKTSKAK--QSIKTSK--DLAPTSKVIKAPT---PK 293
 DB 186 LKPTGTSAKPTTLKPTG---TSKPTTLKPTGTAKTTLKPTGTSKAKPTTLKPT 241
 OY 294 AETTTGPAULTPPKEPTTPPKPEASTPKPEPT---PTIKSAPTTPKPEA---PTTKS 347
 DB 242 EGTAK-----PTTLKPTGTAKTTLKPTLNPTGTSKPTTLKPTGTAKTTLKPTLNPTGTS 297
 OY 348 APTTPKEPAPTTKEPA---PTTPKEPAPTTKEPAPTTKSAPTTPKPEAPTTKKRP 404
 DB 298 AKPTTLKPTGTAKTTLKPTGTSKPTTLKPTGTAKT---PTTLK---PTEGTSKAP 352
 OY 405 TTPKEPAPTTKEPPT--PTTPKEPAPTTKEPAPTTKEPA---PTAKKAPPTTPKEPAP 459
 DB 353 TTPKPTDGTAKPTTLKPTGTSKAPTTTLKPTGTAKTTLKPTGTSKAPTTTLKPTG 412
 OY 460 TTPKEPAPTTKEPPTTPKEPAPTTKSAPTTKPEA---PTTTSAPTTKEPPTPT 516
 DB 413 TTKA---PTTLKPTGTAKT---PTTLKPTGTAKTTLKPTGTSKAPTTTLKPTGPT 466
 OY 517 KEPAPTTKE---PAPTPKPKAPPTTPKEPAPTTTPKEPAPTTTKKAPAPTPKE---PAPT 570
 DB 467 AKPTTLKPTGTAKTTLKPTGTSKAPTTTLKPTGTAKTTLKPTGTAKTTLKPTGTAKT 525
 OY 571 TPKETAPPT--PKKLPPTPEKLAPTT--PEKAPPTPEELAPT--TPEEPTTPPEEPA 624
 DB 526 TPKTKTKTKAKPTTLKPTGTSKAPTTTLKPTGTAKTTLKPTGTSKAPTTTLKPTGPT 585
 OY 625 PTPPKAAPPTPKPEAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKGAPPTTLKPEAPPTP 664

DB 586 TAKPTTLKPTGTSKAPTTTLK-----PTEGTSKAPTTTLKPTGTAKT---PTTLK---PTEG 636
 OY 685 KKPAPKEAPPT-----TTKEPTISTSDKAPPTTPKTAATTPTKPEAPPTPKPEAPPTP 737
 DB 637 TSAQPTTLKPTGTAKTTLKPTGTSKAPTTTLKPTGTSKAPTTTLKPTGTAKTTLKPTG 691
 OY 738 KCTA-----PTTLKPEAPPTPK-----KPAKEAPPTTKGPTSTISDKAPPTPK 783
 DB 692 KPTTLNPTGTSKAPTTTLKPTGTAKTTLKPTGTSKAPTTTLKPTGTAKTTLKPTGTAKT 751
 OY 784 ETAPTPKEPAPPT--PKKAPPTPEPPPTSEVSTPTTK-----EPTTHKSD 834
 DB 752 EGRTTAK---PTTLKPTGTSKAPTTTLKPTGTAKTTLKPTGTSKAPTTTLKPTGTAKT 806
 OY 835 STPELSAETP--KALENSKEGV---PTTKT---PAARKPEMTTAKKTERDLRTPE 888
 DB 807 RT---SAQPTTLKPTGTAKTTLKPTGTSKAPTTTLKPTGTSKAPTTTLKPTGTAKT 858
 OY 889 TTTAAKMTKETATTEKTEESKITATTVSTTODTPPKKITTLKTTTLAPRVTTK 948
 DB 859 GTSKAPTTTLKPTGTAKTTLKPTGTSKAPTTTLKPTGTSKAPTTTLKPTGTAKTTLKPTG 914
 OY 949 K--TTTTIMKPEETAKPKDRATNSKATTPKPKAPK---KPT--STKKKPTMPR 1001
 DB 915 KPTTLKPTGTSKAPTTTLKPTG---GTAKPTTLKPTGTSKAPTTTLKPTGTAKTTLKPT 973
 OY 1002 VRKKTPTT--PRKMTSTMP--ELNPTSRIAEMLOTTTPRNOTNSKLVENPSEDA 1057
 DB 974 TEGSKAPTTTLKPTGTSKAPTTTLKPTGTSKAPTTTLKPTGTSKAPTTTLKPTGTS 1031
 OY 1058 GAGS-ETPHMLLRP 1070
 DB 1032 QANPFTRKRRRP 1045

RESULT 7
 O76894 PRELIMINARY; PRT; 1795 AA.
 AC O76894;
 DT 01-NOV-1998 (TEMBLrel. 08, Created)
 DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE EG:5667.1 OR CG14796.
 GN EG:5667.1 OR CG14796.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton R.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abil J.F., Agbayani A., An H.-J., Andrews-Plannkocn C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Butulis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
 RA Foster C., Gabriellan A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,

[illegible]

Db	597	PSHTTSTTTTAVVSHNRRTISQKTTASITTK-----TTTSKTKTKTIDIPSTTSKL	652
Qy	396	PTTEKKAPATTKEKAPATTKEBPPTTEKAPATTKEKAPATTKEKAPATTAPKAPATTPK	455
Db	653	STTQKTTTTHKFAATSTEKPKTTTEKTSIVSTTKKSTESSPKPSTSTQKPTTTTK	712
Qy	456	EPAPTTPKBPAPTTKEBPPTPKKAPATTKSAPTTKKAPATTGTSAPTTKEBPST	515
Db	713	PSKRTPTTKTKTTTQTITTTTPLASSSTETSTOPEPTTTTPQPTTTTLTVTKTSSTTT	772
Qy	516	TKBPATTKEKAPATTPK-----KAPATTKEKAP-----TTPKBPAP	554
Db	773	TTKEPTTSSPKPTTQKTTSTAPWTGKVALITQKETTPTQSTSTLTFTKTTNNPEPT	832
Qy	555	TTTKKAPATPKKBPATTKEK-ELAPTTPKKLPTTPEKLAPTTEKP-----APTTPE	605
Db	833	STBKPTSTTTPKSTSTKSTVASTSEKTISSPKPTKSTENPTNSVKSALTST	892
Qy	606	ELAPTTPEBPPTTPEBPATTPKAAPNTPKEKAPATTKEBPVT--TPKBPATPKET	663
Db	893	ORATSTTSEPTKT- QNTITTTTPKPTLKSTQCATSTQKSTVITTKKATESSPLTT	951
Qy	664	APTTKKGAPTTTKKAPATTPKKAPKELAPTTKEPT--STTSKDPATTPKGPATTP	721
Db	952	LSTEEPNTPKPLRTPTTTTSTATTTRITTTTTSSTSTSTQKPKSTPTSTRTPT	1011
Qy	722	KEBPATTKEKAPATTPKGATPTTKKBPATTKEKAPATTPKKAPKELAP--TTTKGPTSTSDKPAPT	780
Db	1012	KVTIVISTQNPPTTTSKTSIVT-----TTP-NPSSOTRPTTTTROPSTTASTTSIG	1065
Qy	781	TPKEAPTTKEKAPATTPKKAPATTPTPTPTTSEVSTPTT-TKE-----	824
Db	1066	TTTPTTTPNPNSSSDLTTPVRPCPDSDSKNTACTQELQVNLLELQSPQK	1122
Qy	825	-----PTTIHKRPDSESTPELSAEPKPLENSPKPEPGVPTTKPAAKPEMTTAADKTE	880
Db	1126	EQFTTRHTHTALTSRNTLGGQVPEYMDABSALESQATTTAAPKPSVLAANHLQ	1185
Qy	881	R---DLRTPTTTAAPKMKETATTTTEKTESKITATTTOYSTTQDPTPKITTLKT	937
Db	1186	KLFHISTTPPSREHAP--TORPSSOPSSORSR-GVTIAQMAHNHLATSKFINHSLRL	1242
Qy	938	T-----TLAPKTTTKTKITTTTELMNPEETAPKDKDATSKITTPKQKPTKA-PKPTPS	992
Db	1243	SIQGLASTQKRSIPKPLKLVHNHTTKEED--SEYDSESTEQYTTDEBNELDKTQPRASS	1301
Qy	993	TKKPKTPRVKPKPTPTPKPKMTSTMBELPPTSRIEALQYTTTRPNQTPNSKLEVPK	1052
Db	1302	TTVAAPVPAVSTTTEREPQK-TSSSP--SPT-----KAISSTTTPPIETTTGDELYDSSG	1354
Qy	1053	SED 1055	
Db	1355	SSD 1357	
RESULT	8		
096449			
ID	096449	PRELIMINARY: PRT: 1489 AA.	
AC	096449:		
DT	01-MAY-1999 (TEMBLrel. 10, Created)		
DT	01-MAY-1999 (TEMBLrel. 10, Last sequence update)		
DT	01-MAY-1999 (TEMBLrel. 10, Last annotation update)		
DE	CYST GERMINATION SPECIFIC ACIDIC REPEAT PROTEIN PRECURSOR.		
GN	CAR90.		
OS	Phytophthora infestans (Potato late blight fungus).		
OC	Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;		
OC	Phytophthora.		
OX	NCBI_taxid=4787;		
RA	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=RACE 1-11;		
RA	Goernhardt B.;		
TL	Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.		


```

QY 126 SSSSSSSSSSTIMKISSKNSANBELOKIKLVKDNK-KNPTKK-KP-----TP----- 173
DB 245 STSVKPVSDPSPTKKV-----PVKKEPVEPTPTKNTPKKKKRPMEDEIVEEVK 294
QY 174 KPVYNDAGSGLDNGD-----FVYTPDSTTQHNKVSTSPKTTAKPI-NPRLSPNSD 228
DB 295 EPPVPEKKAPVKKDPAABAAROSPSPKAPKVEPSPVAPPTPVKNPVKKVPPME 354
QY 229 TSKETSLVANKETVEETKTTTNKOTSDGKEKTTSAKKEQOSIKTSKDLAPISKVLA 288
DB 355 VODEPAEEVKKPSAPEKPTPVKKKEPEPSSSTTPSSDPSKAAAVAKROSSPKAPDL 414
QY 289 KATPAE-----TTTGRA-----LTPKEPTPTP-----KEPASTPK 323
DB 415 QADPRAQEVPTPVKNPVKKYKRPWEVDEDEVEEVKQPEAPAKTPTVLKKKEPAKOTA 474
QY 324 EP-----TP-TTISAPPTPKPEAPPTTSAPPTPKPEAPTTTKEPA----- 364
DB 475 KATSKETPEPTPKPPVAPROSSPKVAAPDSAOAPA-TPVKNPVKKWRPWEDEDETRA 533
QY 365 -----PTTPKEAPPTTKEAPPT-----TTKSAPTTKPEP-----APTTKPEAPPTP 407
DB 534 DVNSKPTAKKTPSLAKKOPAPAKESLKPADOTAKAPRDPSPKKVAAPTAPEKTPVLA 593
QY 408 KE-----PAPTPKEPTPTTKEAPPTTKEAPPTPK-EPAPTAAPKAP-----TPKE 456
DB 594 KKEPAGPDSKTEPEKSPKPDPSPKAAVAPKVPKTEVAAVAAKKEPISKPDOTAPKK 653
QY 457 PAPTPKPEAPPTTKEP-----SPTTPKPEAPPTTKSAPTTTKEAPPTTKEAPPTTKEAPPTTKEP 511
DB 654 AEPNSPVVP-PTPVKNPVKKKRPWEDDDAKAVUSLEPEKK-TPVLAAPKAPKPOSEA 711
QY 512 SPTTPKEAPPTTKEP-----EPAPTTKPEAP-----TPKPEAPPTTKEAPPTTKEAPPTTKEAPPTTKEP 566
DB 712 AADVSGSSSKDPLAKKAPVAPKRPDPSPKAVPIKPAKTI-EVPAVAVKKEPVAKSRD 769
QY 567 PAPTPKPEAPPTTKEKILPTTP-----EKLAPTTKEAPPTTKEAPPTTKEAPPTTKEAPPTTKEP 616
DB 770 PSPKAK-AEPNSP--VPTPVKNPVKKKRPWEDDDAPEVAVNVEPEKTPVLA 826
QY 617 PTTKEAPPTTKEAPPTTKEAPPTTKEAPPTTKEAPPTTKEAPPTTKEAPPTTKEAPPTTKEAPPTTKEP 675
DB 827 PVKPRDPSPKAAVAPKSTKTADAPVSVKKREPVSKPEPSPKAEVNSVVP-----PTP 882
QY 676 LKEPAPTPPKAPKELAPPT-TTKEPTSTSDKAPPTTKEAPPTTKEAPPTTKEAPPTTKEAPPTTKEP 734
DB 883 VKNPVKKM-KRPWEDEDEPTEEVAKPSE--PEKKTPTVLAK-KEPEKEXD-APKVAAPKRD 937
QY 735 TTPKSTAPTTLKEBAPT-----TPKKRAP-----KELAPTTTKEGTS-----TTSOK 776
DB 938 PSPKAVPE--KEPAKVAAPKRPDPSPKAPITANQOAPPTPVKNPVKKWRPWEDEDE 995
QY 777 P-----APTPKEP-----APTPKEAPPTTKEAPPTTKEAPPTTKEAPPTTKEAPPTTKEP 825
DB 996 PAEVSAPPEBEKPTPVLAAPKAPKRPD--SPKAAVAAK-PDCKIPEV-PTPVKNPV 1050
QY 826 TTIHKS-----DSTPELSA-EP--TPKALENSPKEPV-----PTTKTA 864
DB 1051 VKMKRPWEDEDEPSEVSAPEPEKKTPTVLAKKAPKTAAPKPDSEAAADVSGSPSKDCK 1110
QY 865 AT-----KEMTTAKDKTTTERDLRTTPEPT--TAAPKTKETATATTEKTESKITAT 916
DB 1111 LSKAPAEKPKPTTDEKDKLSPKAPKPEKAPERAAPKMKRVMDDEDEADDTVA 1170
QY 917 TQVSTTQOTPTPEFKITTLKITTLAKVTTYTKITTTTELMMKPEETAAPKORATNSKAT 976
DB 1171 PSKKPTEDEADPLG-----GPKTDPK-----LKKKAPAEKPTSK----- 1206
QY 977 TPVQAPTAAPKPKTSTKPKTPVAVKPK-----TTPPKMTSTMP 1019
DB 1207 -PAPKEVSEKRPTEPBPV-AAPKMKRPWEDEDEPEADDTMAFPKPDTEP 1259

```

```

RESULT 10
Q9LIE8 ID Q9LIE8 PRELIMINARY; PRT; 1480 AA.
AC Q9LIE8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLDBIA;
RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLDBIA;
RA Pubmed=10907853;
RX Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL; AB001306; BAB03062.1; -.
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 1480 AA; 147153 MW; D1AC0C79F155E732 CRC64;

Query Match 11.2%; Score 817; DB 10; Length 1480;
Best Local Similarity 28.4%; Pred. NO. 5,9e-45;
Matches 343; Conservative 63; Mismatches 563; Indels 240; Gaps 54;

QY 4 KTLPTVLLLLSYEVLYQVYSOELSGKSGFSGKSGCDSDAOCKKYDKCCSPDYESFC 63
DB 6 RNLSEFILLILNFATY-----ARCSPPKPSPPNK--RPKNSV 45
QY 64 AEVNPSPSSSKAPRPSGASQITKSTTKR--SPKRP-----KKTKKVIESSEI 113
DB 46 -----RPKRAVKRPNRK--PTGIKRRPKNNRKKRYUKNNRKRPTKPNRK 98
QY 114 TEENHSENGQESSSSSSSSSTIMKISSKNSANBELOKIKLVK--DNKKNTKKK 171
DB 99 TKPN-----PDKRPPTIKRRPNK-----PRNPKRVNPKNNRKRPTKPN 141
QY 172 TPKPVNDEAGSGLDNGDFKVTPTPTDSTQHNKVSTSPKTTAKPI-NPRLSPNSDTSK 231
DB 142 HKRP-----TKNH-----PDKRPPTIKRRPNK-----PRNPKRVNPKNNRKRPTKPN 175
QY 232 ETSLVNKEETVEKETTNTKOTSDGKEKTTSAKKEQOSIKTSKDLAPISKVLA 289
DB 176 -----PRTNPRESTQPRTNKRRCSPTRPVASPRMATPTTOMPTAT 219
QY 290 PTPKAEITTKGRALTPKEPT--TP--KEPASTTKEPPTTKKAPPTTKEAPPTT 345
DB 220 P-PIKSSVAAPRLATPTATPTTTPVAVPTTTPRIANPRLIMPTATPVAAAPIT 278
QY 346 -----KSAPPTKPEAPPTT-----KEPAPTTKPEAPPTTKEP----- 380
DB 279 NPDISKRPVTP-----PTTPRIAKPRIAMPITSPRAAPRAATPTTTLPRKPRVAT 334
QY 381 -PTTKKAPPTP-KEPAPTTKPEAP--TPPKPEAPPTTKEPPTTKEAPPTTKEAPPTTKEP 436
DB 335 SPIVT--PVTPIAOPRVAPRAATPVAVATPRATPTSPKSPISSTPISSESVATPTAT 392
QY 437 TP-KEAPTAAPKAPPTTKEAPPTTKEAPPTT-----KEPSTTPKPEAPTTTSAPT 491
DB 393 SPKITPRPAKRVATPRLIAKSPITATPTATPVATPTPEKRVAVATPTTPTAT--PP 449

```

```

QY 492 TTKEPAPTTKSAPTPKBP--SPTTKEPAPTTKEPAP--TTPKPAPTTKEPAPT 546
DB 450 VAKPVPETPIAPTPAKPPISPPISPKPPVATPPATPTTTPVKKPPVATPIAP 509
QY 547 TPKEP--APTTPKPPAPAP--KEPAPTPKEPAPT--PKLTP--TTPELAPT 594
DB 510 VAKPVPVTPPTAPPTPIAPPIAKSPVATPTATPPVATPIAPKPPVTPPTTPTATPP 569
QY 595 TPKEPAPTTKEELAPPT--PEEPTTPEEPAPTTPEKAAANT--PKEPAPTTKEPAPT 650
DB 570 VAKPVPATPIAPPTPIAKPPISPPISKSPVATPPATPTTTPPAKPPVATPIATPP 629
QY 651 TPKEPAPTTKEPAP--TTPEKGAFTT--LKEPAPTTKPPAKKELAPTTPEPTSTS 705
DB 630 IAKPVPATPIAPPTPIAPISVAKPPVATPIKTPPAKPPVATPIATPPVAKPPVAT-- 687
QY 706 DKPAPTPKGLAP--TTKEPAPTTKPPAPTTKGAFTPLKKEPAPTTKPPAKKELAPT 764
DB 688 -PPATPPATPIAPPTPPVATPPATPPVATPIAKPPTTI--PPATPPVAMPPIATPP 744
QY 765 TTGPTST--TSKPPAPTTKEPAPT--TPKEPAPTTKPPAPTPP--ETPEPTTSE 815
DB 745 TAKPPIATPIAPVPAKPPVTPPTATPPATPIAPPIAKSPVATPPATPPVATPIAKPP 804
QY 816 VSTPTTKEPTTH--KSPDESTEPESAEPKPALENSPKPEPGVTTKTPAATKPEMT-T 872
DB 805 VATPPTAPPTATPPVAKPPVATPPATPIAPTPPAKPPILTPISKPPVATPPATPTTTP 864
QY 873 TAKOKTTRDRLTPETTTAPKMTKETATTEKTESKINATTTQVSTTTODTPPKI 932
DB 865 PAKPVPATPIATPP--IAKPPVATPPATPIATSPVAKPPVAILPKTPPPAKPPVAIL 922
QY 933 TTKTTTAPKVTYTKKT--TTTEIMKPEETAKKADATNSKATTPKOPKPTAPK 988
DB 923 PIATPPVAKPPVATPPATPIATSPATPPVVT--PPATSPVATPIAKPPTTTP- 978
QY 969 KPTSKKKKKTMPRVKKTPTTPPKMTSTMPELNP--TSRLAE--AML 1032
DB 979 -PTAI-PPVAMPPIATP--PTAKPPVATPIANPPVKKPPVATPIAKPPTVLPPIAKP 1033
QY 1033 OTTTPNOTPNKSL--VEVNPKSEADAGAEGETPMLLRPHVEMPEVTDDMDYLPKPPNO 1090
DB 1034 PVETSPATPTATPPATPPVAILP-----VVKPPVAILPITKPPVATPPVATPP 1078
QY 1091 GIINPMLS 1099
DB 1079 PTAMPPIV 1087

RESULT 11
QSPMO PRELIMINARY: PRT: 1315 AA.
AC QSPMO;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EXTENSIN-LIKE PROTEIN.
GN PEX2.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OC NCBI_Taxid=4577;
OX RN
RP SEQUENCE FROM N.A.
RC TTSSUE-POULEN;
RA Stratford S., Barnes W., Golubiewski A., Cotter R., McCormick S.,
RA Hohorst D., Gao M., Showalter A., Bedinger P.A.;
RT "Pollen Extensin-like (Pex) Genes in a Monocot and a Dicot.";
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: A155297; AAD55980.1;
DR InterPro: IPR001611; LRR
DR InterPro: IPR001998; Xylose_isom.

```

```

DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR003592; LRR-out.
DR Pfam: PF00560; LRR; 3.
DR PRINTS: PR01217; PRICHEXTENS.
DR PROSITE: PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
DR SMART: SM00370; LRR; 3.
DR SEQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;

Query Match 11.2%; Score 816; DB 10; Length 1315;
Best Local Similarity 23.5%; Pred. No. 6,1e-45;
Matches 265; Conservative 106; Mismatches 486; Indels 270; Gaps 35;

QY 32 RFEFSEFEREDCDCAQCKKVKDCCPDYESFCAEVHNFTSPSSK-KAPPSGASQTIKS 90
DB 390 QCAPVLARVEVDCS-----KHCAGAPRTGGGPPSSVPKPPSVGK----- 432
QY 91 TTKRSPPKPKKTKKTVIESEITEHSVSENQESSSSSSSSSTIWKIKSKNSKAN 150
DB 433 --PAAPAPMPTHTPPDVSPPELPPSPV-----PADAPMPTLRSPPAD 476
QY 151 RELQKLLKDKKKKNTKKPKPPVYDEAGSGLDNGDFKVTYPTDSTTOHNKYSTSPK 210
DB 477 EYI-----PTPPVAKSPG-----TSPPARGA-PLQAQAP 508
QY 211 ITTAKPINRPSILPNSDTSKETSLTVNKETTVETKETTNNKQSTDGKEKTTSAKETQ 270
DB 509 AASSPATVVKSSPPAAV----- 527
QY 271 SIKTSAKDLAPTSKVLAKPTKAEITTTGPAITTPKEPTPTTKEPASTTPKEPTPTTI 330
DB 528 -VLPPPAKTPSPAPV-ASPPAPAVSPQPKSPPPAPAVASPPPKSP--PPAPAV 583
QY 331 KSAPTPKEPAPTTKSAFTT-KEPAP-----TTKEPAPTTKEPAPTTTKEPAPT 382
DB 584 ASPPPLMKSPPPAPVASPPQPLKSPPPVLMVSTSVKSPPPVAVASPPPVVSPPL 643
QY 383 TTKSAPTTKE--PAP--TTP--KKPAPTTKEPAPTTTKE--PTTTPKEPAPTTK 431
DB 644 APVSSSPVVKLPPLAPGKSTPPEEKPTPTPVKSSPPPKSLPPTLTTSPPQPK 703
QY 432 EPAPTPKEPAPTPAKKAPATTTKEPAPTTKKEPAPTTTKEPAPTTTKEPAPTTKSAFT 491
DB 704 PTPSTSKPPSPVTLPPKSSPPEEPVSPPOAKSSSP-----PAPVS--SPPP 756
QY 492 TTKEPAPTTKSAFTTKEPSPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 551
DB 757 LKSSPPVPESPPPTPKSSPLAPVSSPQVETSPAPVSSPPPTKSSPLAPVSS 816
QY 552 APTTTPKAPAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 611
DB 817 PPQVETSPAPVSSPPPTPKSSPLAPVSSPQVETSPAPVSSPPPLEBKSSSPS 876
QY 612 PEEPTTPEEPAPTTKAAAPNTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 671
DB 877 SVSPPTTVKSSPPALSSPMTKSSPPAHVSSPPEAKSSPLALISSPSEKSP 936
QY 672 A-----PTTKEPAPTTKPPAKKELAPTTTKEPAPTTTKEPAPTTTKEPAPT 715
DB 937 SSPWVEKTSPPATVSSPPPTPKSSPP--APVSSPPVAVSSPPAPVSSPPPTPKP 993
QY 716 TAPTP-----KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 765
DB 994 LPAPVSSPPPVVKKSSPPPTPVSSPPTPKPLPPPTPVSSPPPTKPPPLPP--APVS 1050
QY 766 TKGPTSTSDKPAFTT--PKETAPTTK--EAPTTTTPKPPAPTTPEPTTSEVSTP 819
DB 1051 SPPPVVKKSSPPAPVSLPPPTPKSPPTRVSSRPVVKCCPPATLVSSPPAPKSLPP 1110
QY 820 TTTKEPTTIHKSPESTPELSAEPKPALENSPKPEPGVTTKTPAATKPEMTTAKDKTT 879
DB 1111 TPVSSP-----PPEVAKSSPPPTPVSSPPAPKSSPPPTPVSSPP----- 1150

```

```

OY 880 ERDLRTP-----ETTAPKMKETATITTEKTESKITATTTQVSTTTODTTPKITT 934
DB 1151 --ELKSSPPAPVSSPPSAKSSPPAPVSLPPEPEKSSPPAPISPPPAKSP----- 1203
OY 935 LKTTTLAPKVTYTKKITTTEIMNKPEETAKPKDRAITNSKATTPKQKTKAPKPTSTK 994
DB 1204 -----PPPA-----PMSSLPPPVKSSPPPAVSSPPPMKSSPPPAIS 1242
OY 995 KRTMPKRVKPKTTPTPRKMTSTMPBLNPTSRIAEAMLOTTTRPNOTPKSLVEVNPKE 1054
DB 1243 SPPAP--VKPPSLPPAPVSSPPAV-----TSAP-----PKKE 1275
OY 1055 DAGGAGETPRHMLRPHVMEVTDMDLPRVNOGITINMLSDE 1101
DB 1276 EOSTA-----PPAELPPSPFNDIILPIMANK 1303

RESULT 12
P70670 PRELIMINARY: PRT: 2187 AA.
AC P70670:
DB 01-FEB-1997 (TREMBLrel. 02, Created)
DB 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DB 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
MUSCLE-SPECIFIC FORM GP220).
GN NAC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96312450; PubMed=8698236;
RX Yotsof W.V., St Arnaud R.;
RT "Differential splicing in of a proline-rich exon converts alphaNAC
RT into a muscle-specific transcription factor.";
RL Genes Dev. 10:1763-1772(1996).
DR EMBL: U48364; AAB18734.1; -
DR EMBL: U48363; AAB18732.1; -
DR MGD: MGI:106095; Nac.
DR InterPro: IPR002715; NAC.
DR InterPro: IPR003037; TS-N.
DR Pfam: PF01849; NAC; 1.
DR Pfam: PF02094; TS-N; 1.
SQ SEQUENCE 2187 AA; 220599 MW; 003646A864DEBD CRC64;

Query Match 11.2%; Score 812; DB 11; Length 2187;
Best Local Similarity 27.4%; Pred. No. 1.9e-44;
Matches 311; Conservative 128; Mismatches 449; Indels 246; Gaps 51;

OY 68 NPTSPSSKAPPSGASQITKTRSPKPKKTKKYIESEETTEHSEVSENOESS 127
DB 746 SPT-PPSSKAPVPSTGA-----PPSPKAPIVPTSSISISQVPAEILPSPQ 792
OY 128 SSSSSSSSTIMKISSKSNANRELQKLVKNKNTKPKPTKPPVDEAGSGLDN 187
DB 793 KTEVNTASRLISAVOSPKVPIKSD-----VTPSPKTKSATAVAK----- 833
OY 188 GDFKVTTPDSTQHNK-----VTSPTKITTAKIPRPSLSPNSDTSKETSLS--TVNKE 240
DB 834 -----DTSATLSLKSVPATVTSLSPPKAPVASEATVPLEITPSLKNLAAATPK 885
OY 241 T-----TVEKTEYTTNKTSTGKEKETSASKEIOSIKTSKADLAPTSKVLAKPTPKA 294
DB 886 TLATSIPTKVPSPQKTPKSVSLKAPAMTSKAT---ELASKDVSSQ--FPEVPL 940
OY 295 ETTTKPALITPREPTTPKPEASTTPKEPTTITKSAPTTPKE--PATTTKSAPTTPK 353
DB 941 QHV---PTTSPKSPVSDLSGALTSPPKGPATLAETPTVYKPKSPKPAASKKTPATPS 997
OY 354 -----EPAPTTKKEPATTPKPEPATTTKPEPATTT--TKSAPT----- 389

```

```

DB 998 PEGVTAVLEIPPSKAPKTAAPKESASSSSKRAKTAIVSKKEIPSGVTAVLEISLP 1057
OY 390 -----TPKEPATTPPK--PATTPKEPATTPKEPTTPTPKEPATTPKEPATTPKE 440
DB 1058 LKETSASLPEKASSPKSPKAGPK-----TPPGVTAVPEISLIPKETQONATPNE 1114
OY 441 PAPAPKPKPATTT--PKEPAP-----TPKKEPATTPKEPSPTTPK-----EPAPTTKSA 489
DB 1115 SLAASSQKRSKTSVPEKTEPGVTAVLEIPSAQKAPKATVAPKQIPPEDEAVTILAGS 1174
OY 490 PTTKEPATTT--KSATTPK-----PPSPTTKEPATTPKEPATTPKPKPATTP-- 539
DB 1175 PLSPKKSKTAARKEAPATSVGVIAVSGEISPSPKTSKTAAPKENSATLPPKRSPTKA 1234
OY 540 -PKE-----PATTPKEPAP-----TTTKRPATPAPKEPATTPKEPATTT 579
DB 1235 APKETPATSEGVTAVSEISPSPPKSGVAVTILPGAPNALAE--SPASKKVPKTA 1293
OY 580 PKKLTPTTPKLAPTTPK--PATTPKEPATTPKEPTTP-----PEEPATTPKA 630
DB 1294 APESTSTP-----SPQIKVAGKPEASATPPSKKTEKTAAPKETSAPSEGVTAVLE 1347
OY 631 AAPTPEPATTPKPE--PATTPKEPATTPKETAPTTKGTAPTTLKKEPATTP----- 684
DB 1348 IPPSPKAPKTAAPKETPAP--PEGATTAPOIIPSPKSKKAGSKE--TPTTSPSECV 1404
OY 685 -----KKPAKELAPTTKEPTS-----TSDKPAPTTPKGTAPTPKE 723
DB 1405 TAAPLEIPSSKKTSSKMASPKETLVPSSKLSQVGPKEITLEGATVPLEIPSSHKKA 1464
OY 724 PATTPKE--PATTPKGTAPTLKEPATTPPKKPAKELAPTTTGGPTSTSDKAPATTP 782
DB 1465 PKVDDKQVPLPSPK--DAPTTLAE--SPSSPKK--APKTAAPSER--VTVPEKPA--TPQ 1519
OY 783 KETAPTPKEPATTPPKKAPATTPETP-----PPTSEVSTPTTK-----EPTTKHS 831
DB 1520 KASGTASKVPVPAETQEVAVASRETPVPAVPKKNSSHKTKITELKEAPATLPPS 1579
OY 832 PDESPELSAEPPTKALESPKE--PGVPTTKPAATKPEMTTANDKTTERDLRTTPEPT 890
DB 1580 PTKSPKIPSSKAPRT--SAPKEFAPSISIK--PVTT--SLAQTAAPSLQKAPSTTIPREN 1634
OY 891 TAAPKMKETATTTTEKTESKITATTTQVSTTTODTTPKITTTLAPKVTYTKT 950
DB 1635 LAAPAV---LPVSSKSPAAAPARASASLSPATAAPQTAAPKETAITTPSCKKAATETPIET 1690
OY 951 ITTTEIMNKPEETAKPKDRAITNSKATTPKPKTPAKPKPTSTKKPKTMPVRKPKTTPPT 1010
DB 1691 STAPSLGAPKETSE-----TSVSKVLMSSP-----PKKASSSKRASLTP----- 1730
OY 1011 PKMTSTMPBLNPTSRIAEAMLOTTTRPNOTPKSLVEVNPKESEDAGAGETP 1064
DB 1731 -----ATILPSLKEASVLS-----PTATSSGKDSHISPV--DACSTGTTPP 1770

RESULT 13
O9VEL9 PRELIMINARY: PRT: 2112 AA.
AC O9VEL9:
DB 01-MAY-2000 (TREMBLrel. 13, Created)
DB 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DB 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG4090 PROTEIN.
GN CG4090.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;

```

RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanosides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Azpil J.F., Agbayan A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottler P.,
 RA Butris K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Domres M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dudin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Guan P., Harris M.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Huan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshneft A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relberte K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitslks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissensbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster.";
 RA Science 287:2185-2195(2000).
 RA EMBL: AE003716; AAF55402.1;
 RA Flybase: FBgn0038492; CG4090.
 RA Interpro: IPR002557; Chitin_binding.
 RA Interpro: IPR000561; EGF-like.
 RA Pfam: PF01607; Chitin_bind_2; 10.
 RA SMART: SM00494; ChIBD2; 11.
 RA PROSITE: PS00022; EGF_1; 1.
 RA DR PROSITE: 2112 AA; 219547 MW; B91018E551A5D36 CRC64;
 RA SEQUENCE

Query Match 10.8%; Score 786.5; DB 5; Length 2112;
 Best Local Similarity 26.9%; Pred. No. 8.2e-43;
 Matches 313; Conservative 108; Mismatches 380; Indels 361; Gaps 49;
 21 QVSSQELSCGRCFESFERECDCD-----AOCKKYDKCCPDYSEFCAEVHNPTSP 73
 953 QSSSSNNSTSTQKPKRKECESETFLADNENCSKFRVCDNKGFTYV-SFTCP 1011
 74 SSKRAPPASGASQTIKSTTKRSPKPPNKKTKTKYI-----ESEITIEHVSSESSSS 129
 1012 NTLMDFPANSNCNHPDQIOTK-----PLKCKRVVSOGSSSSNSTSSSSSSSSN 1064
 130 SSSSSSTIKIKISSKNSANRELÖKKLKYDKKKNRKTKKPPKPPVVDAGSGLDND 189
 1065 SSSSSS-----SSNS-----GSSSNTGS 1084
 190 FKVTPTSTQHNKVSSTPKITTAAPINRPSLPNSDTSKTSLTVNKKETIVETKETT 249
 1085 SSNSGASSSGSSNOGSSS-----NSGSSSSNSGNSGNTS-----SS 1122
 250 TTNKQSTQDKEKTKSAKKEQSIKETSAT-DIAPTSKVLAKPPKPAETTK----- 299
 1123 TSSSSSSNNNNNOGSSSSSSSSSSSSSKPRPSETCKYVNGQFIGDRDCAKFRVCDNR 1182

300 -----GP-----ALITPKPEPTPKPEASTPEKPTPT 328
 1183 GGFNNWPSGPGTVDADQMOACNHMAVKEGGIAPPTSTPTSR-PTTASTSRSDQ 1241
 329 TTKSAPTPKPEAPPTTTSAPTTKEPAPTTTPKEPAPTTKEPAPTT----- 375
 1242 TSTSRPTG-----PTTAPRYNARPTTSSPTTASSQTTSPVQAQPNNDGKCRSGFMADP 1297
 376 -----TKEP-----APT-----TTKSAPTPKPEAPPT 398
 1298 NNCSEKRCVNRNKGSTIFQCGAGTVWDQDLOTCHNHNKNCSTGSESTTPKPC--E 1355
 399 PKKPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPA 458
 1356 PATNGTANSTSTTP-----PTTDLPLPTSTGLP-PTTTELPPT-----TTDLPLP 1405
 459 PTPKPEAPPTTKEPSPTPKPEAPPTTKSATTTKEPAPTTKEPAPTTKEPAPTTKE 515
 1406 TTTTRLPPTTTSILPPTTTGHPPTTTGADPTTTLSEETSTVTTSPESTTTPSTT 1465
 516 TKRPAP-----TPPK-----E 526
 1466 TMRPLDAGTECTGEGYADPEDCRKYRCINAGASRYKNFTCPKGTGNNERVQCDYVE 1525
 527 PAB-----TPPKPAPTPP-----KEPAPTTKE-PAPTTTPKPAVAPKEPAPTTKEPAT 578
 1526 NTRCKSLAPRPTTTPSESKDPGSTTPQSDPEPTTVKPI-TKPTL-EPSTEROKPT 1583
 579 TPKKLPTTPEKAPPTTPEKAPPTTPELAPTTPEPTTPEDEAPPTTPEKAAANTPE 638
 1584 TOYPERKPTTEE--PEKPOKPT-TTEYPOKPTTEEPTEKOKPI-TTEYPOKPTTEE 1639
 639 PAPTPEKAPPT-PEKAPPTTKEPA-----PTPK-GLAPTTKEPAPTTTPKPAV 690
 1640 PTEPKPKPTTTEPKPTTEETTSIPQYNPTTSVQYNPTTPIVEETTSIPGK 1699
 691 ELAPTTKEPTSTSDKPAATTTKGAATTPKPAATTPKE-PAPTPKGATPTLKEPA 749
 1700 ---PTTGGPTTTT-TPSTTDAIGPTTSKPEPTTTESSPESTPECSVTTLQPEQ 1755
 750 PTPPK-----PAKELA-----PTTKGPTSTSDKPAATTPKPAATTPKEPATTP 798
 1756 PNVCSEGPFPDEDSRYRCVDAKNGKYVAFKCGKGTWMDSTETCNVADOVSG 1815
 799 KKPAPTPPEPTTTSVSPPTTKEPTTHKSPDESPTPLSAPPLKALENPKPEGPV 858
 1816 N-----CSSGQTTTPGTTTEPGTTESPTSSGKP-----ETTSKAPEN----- 1852
 859 TTKPAATKPEMTTAKDKTTERDLPTTPETTAAAPKMTKEATTEKTESKITATTTQ 918
 1853 -TTTWA--PETTTT-----SSPETTTT--VASETTTTSCT-----TTTA 1887
 919 VTSSTTQDTTFFTKTTLTKTTLAPKVTYTKTITTEIMNKKPEETAKPKDRAATNSATTP 978
 1888 TPETTTKPKP-----ETTIIAGEETSTSKSPYTE-----SP 1920
 979 KPQKPTKAPKPTSTKPKTMP 1000
 1921 APSTNTSAP-----CPETGP 1935
 RESULT 14
 Q41805 PRELIMINARY; PRT; 1188 AA.
 AC Q41805;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE EXTENSIN-LIKE PROTEIN PRECURSOR.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC clade;
 OC Panicoidae; Andropogoneae; Zea.

OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=873; TISSUE=POLLEN;
 RA Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 234455; GenBank: 2346; J1346.
 DR Mendel: 14346; Zema: 2368; 14346.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 3.
 DR SMART: SM00370; LRR; 4.
 DR Signal.
 KW Signal.
 FT SIGNAL.
 SQ SEQUENCE 1188 AA; 120980 MW; 2C77C7F8D7130149 CRC64;

Query Match 10.8%; Score 784; DB 10; Length 1188.
 Best Local Similarity 29.9%; Pred. No. 6.6e-43;
 Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

OY 171 PPPKPPVVDAGSGLDNGDFKVTPTDTSTQHNKYSTPKITTAKPINRPSL-----PPN 226
 DB 460 PPHSPPAD-----DYVPTPPVPGKSPSPATSPBQVQPAASSTPSLVKLSPPQ 510
 OY 227 SDTSKETSIVNKEETVEKETTNNKOTSDGKEKTSKAKTOSIEKTSKADLAPTS 286
 DB 511 AVVG-----SPPPVKTSSPPATG-----SPSPPPVSVV 541
 OY 287 -----LANPTKAEKTTTGKPALEPTTPKKEPTTPKKEAPTTPKKEAPTTPKKEPA 341
 DB 542 SPPPVKSPPPAPVGSPPPEKSPPPAPVASSPPPVKSP--DEPTLVASPPPVKSP 599
 OY 342 PTTTSAATTP--KRAPTTTKEAPATTKEAPATTKEAPATTKEAPATTKEAPATT 398
 DB 600 PPAPVSPPPVKSPPPTPVASPPPPVASSPPPMKSPPPPTPVSSPPPEKSPPP 659
 OY 399 PKKPAPTTPKKEAPATT-----PKE---PTPTTKEAPATTKEAPATTTPKKEAPATAK 449
 DB 660 PAKSPPEEKPPTPTSVKSSPPPEKSLPPTLLPSPPQEKPTPTSPK--PSSDEK 718
 OY 450 APTTPKEAPATTKEAPATTKEPSPTTPKKEAPATTKSAPTTTKEAPATTKSAPT 509
 DB 719 SP--PKEVSSPPQTP-----KSSPPAPVSSPPPTPVSSPPALAVSSPSVSKSP----- 768
 OY 510 EPSPTTKEAPATTKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAP 569
 DB 769 PPALSSPPAPVQKSSPPPVQVSSP--PPAKSSPPAPVSSPPQVKEKTSPPAPLSSP 827
 OY 570 TTPKETAP---TTP-----KLTPTTPEKLAPTTPKEKPA-----TTPPELAPTTPEPT 616
 DB 828 LAPKSSPPHVVSPPPVVKSPPPPAPVSSPPLTPKASPPAHVSSPPEVVKSTP--PA 885
 OY 617 PTPEEAPATTTPKKAANTPE--PAPTTPKEAPATTTPKEAPATTTPKETAATTTPKETA 675
 DB 886 PTTTIS--PSEKSSPPPTVSLPPTIVKSSPPAMVSSP--PMTKSSPPPVVSSPPPT 943
 OY 676 LKE-----PTTPK--PAKELAPTTKE--PTSTSDKAPATTTPKGTAPTT 721
 DB 944 VKSSPPAPVSSPPATKSSPPAPVNLPPPEVKSPPPVVSSPPA--PKSSPPAP 1000
 OY 722 -KEAPATTTPKEAPATTTPKGTAPTTLKEAPATTTPKKAAPLATTTPKGTSTSDKAP 780
 DB 1001 MSSPPPEVSSPPPPAVSSPPPVKSSPPAPVSSP--PPVKSPPPPAPVSSPPPV 1057
 OY 781 -TPKETAPTTTPKKAAPTTTPKKAAPTTTPPTTSEVSTPTTKEPTTHKSPDESTPEL 839
 DB 1058 KSPPPAPVSSPPPVKSSPPPVAVS--SPPPVKSSPPAPVSSPPPKSPPPPAPV 1115
 OY 840 SAAPTTPALENSPKKEGVPTTKPATKEMTTAKDKTTENDLRRTPT 889
 DB 1116 SPPAP-----VKPSP--PPAPVSSPPPVVTPAPKKEOSLPPPAS 1158

RESULT 15
 O9XDH2 PRELIMINARY; PRT; 763 AA.
 ID O9XDH2
 AC O9XDH2:
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE PROLINE-RICH MUCIN HOMOLOG.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RA Espitia C., Lacleite J.L., Mondragon-Palomino M., Martens A.,
 RA Zhang Y., Moreno C., Singh M.;
 RT "Cloning and characterization of a new member of the PGSR family that
 is a useful marker of polymorphism in Mycobacterium tuberculosis."
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF071081; AD41594.1;
 DR InterPro: IPR002951; Atrophin.
 DR InterPro: IPR002965; P-rich_extensin.
 DR InterPro: IPR003882; P-rich_extensin.
 DR PRINTS: PR01222; ATROPHIN.
 DR PRINTS: PR01217; PRICHEXTENSIN.
 DR PRINTS: PR01218; PSTLEXTENSIN.
 SQ SEQUENCE 763 AA; 75034 MW; 39168EC45A5916F8 CRC64;

Query Match 9.8%; Score 715; DB 2; Length 763;
 Best Local Similarity 28.2%; Pred. No. 1.3e-38;
 Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

OY 290 PTPKAEKTTGKPALEPTTPKKEPTTPKKEAPTTPKKEPTTPKKEAPTTPKKEAPTTPK 349
 DB 3 PVP-----APRALALPAP 46
 OY 350 TTPKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAP 409
 DB 47 PCD--PAP-----PAPPKKSKAPPPVPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 490
 OY 410 PAPTTPKEPTTPK-----PKEAPTTPKEAPATTTPK--EPAPATK--KRAPTTTPKEAPATT 463
 DB 91 SRALPBCPPPVVLPDEPPAPAPPPAPNSPPPEPPAPKPPVAPVP--PVPNSPP 148
 OY 464 EP--APTTPKEPSPTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTP 521
 DB 149 FPPPPALNPPAP-----PAPPLANSPLLPAPAPAPAGT--PPAAPPPVPPAPAKSKPA 201
 OY 522 TTPKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAPATTTPKKEAP 577
 DB 202 SPPPPAP-----PPAPATMEPPPLVPVPDIPKSTPPAPAPAPAPAPAPAPAPAPAP 256
 OY 578 TTPKKLTPPTPKLAP-----TTPKEAPATTTPKEELAPTTPEPTTPPEEAPATT-- 627
 DB 257 PVPNKIIPAPP--APVAVAAVLVAPCPPLPLPNNHPPAPAPAPVGVPLAPLPSHP 313
 OY 628 ---PKAAPNTTPKEAP----- 641
 DB 314 PAPSPAPVGVPLAPLPLSGRVSVMKGSFTTLSTFCRCVSGEVLALNPSRPSPL 373
 OY 642 -TTPKEAPATTTPKKEAPATTTPKETA-----PTTPKGTAT--TTTKEAPATTTP--KKAPKE 691
 DB 374 TTTTPALAPAPL 431
 OY 652 LAPTTTKEPTTSDKAPATTTPKGTAPTTTPKKEAP-----PTTPKEAPATTTPKGTAT 745
 DB 432 --PDPGKFWTTPPLAPAPPEPK--TVPVLPDGGSCPPSEKPPNPPAPPEPEPKSSAPLPP 488
 OY 746 KEAPATTTP--KRAPKELATTTPKGT--TSTTSKAPATTTPKET--APTTPKEAPATTTP 798

```

Db 489 APPAPMSAVRVPSPPIPPAPAPRASMPALPPAPSPPATRLCPPLPSPAPANSF 548
QY 799 KKPAPPTTPTTPTTSEVSTPTTKEPTTIHKSP-----DESTPELSAEPYPKA--LENSP 852
    ||| ||| | : | : | : ||| ||| : |||
Db 549 --PAPAPPTPPKLS--ANPCPPVPPAPBNRPPAPAPAPPELPAPDPPTPPVANSF 604
QY 853 KPCVPTTKPPAATKPEMTTAKDKTTERDLRTTPTTAPKMKETATTEKTESKI 912
    | | : | : | : ||| | : | :
Db 605 --PAPAPAPPSALPFINPPA-----PPTPAAPK-----SRPAL 637
QY 913 TATTQVSTTTODTPEPKITITLKTTLADPKVTTKKTITTEIMNKPEETAKPKDRATN 972
    | : ||| | : | : | : | : | : | :
Db 638 PAAPAPAPAPVVRATP-----PPAPAPAPPN 665
QY 973 SKATTPKOKP---TKAPKPTSTKKKMPRYRKPKTTPTRKMTSTM-----PELNP 1023
    | | | | | : | : | : | : | : | : | :
Db 666 SMALPPAPDPPIPLATPPAPAPPLPMSPPAPPLPPAADPPAPPLTINOPSPPLAP 725
QY 1024 TSRTAEMLOTTTRPNQTPNSKLVEVNPKSEDAAGAGET 1063
    | | | : | : | : | : | : | :
Db 726 VPGAPLAPLPINGRVPFARKNSLI-----GSSSGDT 756

```

Search completed: April 26, 2002, 16:29:14
 Job time: 569 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:29:14 ; Search time 210.7 Seconds

(without alignments)
881.660 Million cell updates/sec

Title: AA3
Perfect score: 6814
Sequence: 1 MAMKTLPIYLILLLSVFVIQ.....ARAITRRSGOTLSKWNCP 1270

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: SP_Archaea:*
3: SP_Bacteria:*
4: SP_Fungi:*
5: SP_Human:*
6: SP_Invertebrate:*
7: SP_Mammal:*
8: SP_Mhc:*
9: SP_Organelle:*
10: SP_Phage:*
11: SP_Plant:*
12: SP_Rodent:*
13: SP_Virus:*
14: SP_Vertebrate:*
SP_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6727	98.7	1404	4	Q92954
2	6712	98.5	1404	4	Q92954
3	3369	49.4	1054	11	Q9BM99
4	1713.5	25.1	401	6	077765
5	971.5	14.3	1079	5	Q9N4S7
6	936.5	13.7	1049	5	Q9N4S7
7	865.5	12.7	1795	5	076894
8	855	12.5	1489	10	Q96449
9	811.5	11.9	1315	10	Q9SPM0
10	802	11.8	1274	5	Q20007
11	800.5	11.7	1480	10	Q9LIE8
12	784	11.5	1188	10	Q41805
13	776.5	11.4	2187	11	P70670
14	756	11.1	2112	5	Q9VEL9
15	745	10.5	763	2	Q9VDH2
16	707.5	10.4	555	10	Q9FPO6
17	671	9.8	1151	13	057580
18	670	9.8	1229	5	Q9A185
19	668	9.8	2284	5	Q9VPG1

20	667.5	9.8	4880	11	Q9LJL1	Q911c1 rattus norv
21	667.5	9.8	5085	11	Q9JRS6	Q91ks6 rattus norv
22	667	9.8	6677	5	Q9N435	Q9N435 caenorhabd
23	662	9.7	2089	4	Q14676	Q14676 homo sapien
24	653.5	9.6	1514	5	Q9GGM7	Q9GGM7 leishmania
25	651.5	9.6	3570	4	Q99552	Q99552 homo sapien
26	647	9.5	4833	11	Q90YX6	Q90YX6 mus musculu
27	647	9.5	5038	11	Q90YX7	Q90YX7 mus musculu
28	642.5	9.4	1612	5	Q9VYQ2	Q9VYQ2 drosophila
29	641.5	9.4	3507	5	Q23587	Q23587 caenorhabd
30	635.5	9.3	7962	4	Q10465	Q10465 homo sapien
31	635	9.3	2768	5	Q9VC00	Q9VC00 drosophila
32	633	9.3	489	10	Q41707	Q41707 vigna unguis
33	632	9.3	761	10	Q9ZQ10	Q9ZQ10 arbidopsin
34	632	9.3	6632	5	Q17362	Q17362 caenorhabd
35	630.5	9.3	990	13	Q91803	Q91803 xenopus lae
36	626.5	9.2	971	5	Q9XVSA	Q9XVSA caenorhabd
37	625	9.2	6642	5	Q01761	Q01761 caenorhabd
38	622.5	9.1	839	2	Q9RX57	Q9RX57 delinococcus
39	615	9.0	234	5	Q9N3Y8	Q9N3Y8 caenorhabd
40	608	8.9	801	5	Q23635	Q23635 caenorhabd
41	607.5	8.9	924	12	Q99307	Q99307 epstein-bar
42	605	8.9	379	5	Q27929	Q27929 drosophila
43	600.5	8.8	1893	5	Q9NKC9	Q9NKC9 drosophila
44	598.5	8.8	409	10	Q9SBM1	Q9sbm1 volvox cart
45	577	8.5	956	10	Q9LJ64	Q91j64 arbidopsin

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	1404 AA.
Q92954	Q92954	PRELIMINARY:	PRT:	1404 AA.
AC	Q92954	PRELIMINARY:	PRT:	1404 AA.
DT	01-FEB-1997 (TREMBLrel. 02, Created)			
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	MEGAKARYOCYTE STIMULATING FACTOR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
RA	Kellerer K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,			
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
RL	Blood 78:279-279(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,			
RA	Fitzgerald M., Scalitro J., Kellerer K., Preissner K., Kriz R.,			
RA	Jacobs K., Turner K.,			
RL	(in) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,			
RL	Mosher D.F. (eds.);			
RL	Biology of vitronectins and their receptors., pp.45-52,			
RN	Elsevier Science Publishers B.V. (1993).			
RP	SEQUENCE FROM N.A.			
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
RA	Kellerer K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,			
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.,			
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.			
RN	EMBL, U70136; AAB09089.1; -			
DR	InterPro: IPR000585; Hemopexin.			
DR	InterPro: IPR001212; Somatomedin_B.			
DR	InterPro: IPR002400; GF_cysknot.			
DR	Pfam: PF00045; hemopexin_2.			
DR	Pfam: PF01033; Somatomedin_B; 2.			
DR	PRINTS: PR00438; GFCYSKNOT.			

PRINTS: PRO0022; SOMATOMEDINB.
 DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
 DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
 DR SMART; SM00120; HX; 2.
 DR SMART; SM00201; SO; 2.
 SO SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

Query Match 98.7%; Score 6727; DB 4; Length 1404;
 Best Local Similarity 90.3%; Pred. No. 0;
 Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

```

QY 1 MAMKTLPIYLILLLSVFIQVSSO----- 25
Db 1 MAMKTLPIYLILLLSVFIQVSSODLSSCAGRCGEYSRDATNCNDYNGOHMECCPDF 60
QY 26 -----ELSCGRCGFESFEGREDCDAQCCKKYDKCCPDYESFCAE----- 65
Db 61 KRVCTAELSCGRCGFESFEGREDCDAQCCKKYDKCCPDYESFCAEVHNPTSPSSKAP 120
QY 66 ----- 65
Db 121 PPGASQTIKSTKRSPKPKKTKKVIKIESEITEHSHSVENQSSSSSSSSSTIW 180
QY 66 -----VQDNKKNRTKKKPTPKPVYVDAGSGLDNGDFKVTTPDTST 106
Db 181 KIKSSKNSANBELOKKLKVKNKKNFKKKPTPKPVYVDAGSGLDNGDFKVTTPDTST 240
QY 107 TOHNKVSSTPKITTAKEINPRPSLPNSDTSKETSLVYKETTVEKETTNNQTSIDG 166
Db 241 TOHNKVSSTPKITTAKEINPRPSLPNSDTSKETSLVYKETTVEKETTNNQTSIDG 300
QY 167 KEKTSAAKETOSIEKTSADLAPTSXYLAKPTPKAETTTKGPALTTKEPTTPKEPAS 226
Db 301 KEKTSAAKETOSIEKTSADLAPTSXYLAKPTPKAETTTKGPALTTKEPTTPKEPAS 360
QY 227 TTPKEPTPTTKISAPTTKEKAPATTTSAPTTKKEPAPTTTKEPAPTTKKEPAPTTKEP 286
Db 361 TTPKEPTPTTKISAPTTKEKAPATTTSAPTTKKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
QY 287 APTTKSAPTTKEKAPATTTPKKAPATTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 346
Db 421 APTTKSAPTTKEKAPATTTPKKAPATTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 480
QY 347 EPAPTAAPKAPATTTPKEPAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 406
Db 481 EPAPTAAPKAPATTTPKEPAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 540
QY 407 TTKSAPTTKEKESPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 466
Db 541 TTKSAPTTKEKESPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 600
QY 467 APTAPKEPAPTTPKETAPTTPKKLTPPTPEKLAPTTPEKAPATTPEDELAPTTPEEPTPT 526
Db 601 APTAPKEPAPTTPKETAPTTPKKLTPPTPEKLAPTTPEKAPATTPEDELAPTTPEEPTPT 660
QY 527 PEPEPAPTTKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 586
Db 661 PEPEPAPTTKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 720
QY 587 APTTPKPKAPKELAPTTTKEPTSTSDKAPATTPKGTAFTTPKEPAPTTTPKEPAPTTPKG 646
Db 721 APTTPKPKAPKELAPTTTKEPTSTSDKAPATTPKGTAFTTPKEPAPTTTPKEPAPTTPKG 780
QY 647 TAPTTKEPAPTTTPKKAPKELAPTTTKGPTSTSDKAPATTTPKEPAPTTTPKEPAPTTPK 706
Db 781 TAPTTKEPAPTTTPKKAPKELAPTTTKGPTSTSDKAPATTTPKEPAPTTTPKEPAPTTPK 840
QY 707 KRPAPTPPEPTPTSEVSTPTTKEPTTHKSPDESTPELSAEPYKALENSKEGAVPT 766
Db 841 KRPAPTPPEPTPTSEVSTPTTKEPTTHKSPDESTPELSAEPYKALENSKEGAVPT 900
QY 767 TKTPAATKPEMTTAAKDKTERDLRTTPTTTAAPKMKETATTTTEKTESKITATTTTQY 826

```

```

Db 901 TKTPAATKPEMTTAAKDKTERDLRTTPTTTAAPKMKETATTTTEKTESKITATTTTQY 960
QY 827 TSTTODTTPFKITTLTKTTTLTAPKYTTTKKITTITEIMNKPDEETAKPRDRATNSKAITPK 886
Db 961 TSTTODTTPFKITTLTKTTTLTAPKYTTTKKITTITEIMNKPDEETAKPRDRATNSKAITPK 1020
QY 887 POKPTAPKPKSTYKPKTKTPVYRKKTTPTRKMTSTMPLEINPTSRITAEMLQTTTRPN 946
Db 1021 POKPTAPKPKSTYKPKTKTPVYRKKTTPTRKMTSTMPLEINPTSRITAEMLQTTTRPN 1080
QY 947 QPNSKLVVWVKSEPDAGAGEPTPHMLLRPHVFMPEVTPDMDYLRVYNOGIIINPMIS 1006
Db 1081 QPNSKLVVWVKSEPDAGAGEPTPHMLLRPHVFMPEVTPDMDYLRVYNOGIIINPMIS 1140
QY 1007 DETNIGNKRPVDGLTLTRNGTLVAERGHYFMMLSPSPSPSPARRITTEWGLSPIDYFT 1066
Db 1141 DETNIGNKRPVDGLTLTRNGTLVAERGHYFMMLSPSPSPSPARRITTEWGLSPIDYFT 1200
QY 1067 RCNCEKTEFFFKDSQYWRFTNDIKDAGYFKPIFGGGLTGQIVAAALSTAKYKNMPEVY 1126
Db 1201 RCNCEKTEFFFKDSQYWRFTNDIKDAGYFKPIFGGGLTGQIVAAALSTAKYKNMPEVY 1260
QY 1127 FFKRGSIQOYIKQBPVOKCPGRRPALNYPYGEWTOVRRRRFERAIGPSQTHTRIOY 1186
Db 1261 FFKRGSIQOYIKQBPVOKCPGRRPALNYPYGEWTOVRRRRFERAIGPSQTHTRIOY 1320
QY 1187 SPARLAVODKGVLANEYKYSILMRGIPNVYTSALSLPNTRKPDGVOYVAFSDQYNNIDV 1246
Db 1321 SPARLAVODKGVLANEYKYSILMRGIPNVYTSALSLPNTRKPDGVOYVAFSDQYNNIDV 1380
QY 1247 PSRTARATITRSGOTLSKWYNCP 1270
Db 1381 PSRTARATITRSGOTLSKWYNCP 1404

RESULT 2
Q9BX49 PRELIMINARY; PRT; 1404 AA.
ID Q9BX49
AC Q9BX49:
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE BG17416.2 (MSF: MEGAKARYOCYTE STIMULATING FACTOR ).
GN BG17416.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wray P.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133553; CAC36090.1; -.
SQ SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEB5 CRC64;

Query Match 98.5%; Score 6712; DB 4; Length 1404;
Best Local Similarity 90.2%; Pred. No. 0;
Matches 1267; Conservative 0; Mismatches 3; Indels 134; Gaps 2;

QY 1 MAMKTLPIYLILLLSVFIQVSSO----- 25
Db 1 MAMKTLPIYLILLLSVFIQVSSODLSSCAGRCGEYSRDATNCNDYNGOHMECCPDF 60
QY 26 -----ELSCGRCGFESFEGREDCDAQCCKKYDKCCPDYESFCAE----- 65
Db 61 KRVCTAELSCGRCGFESFEGREDCDAQCCKKYDKCCPDYESFCAEVHNPTSPSSKAP 120
QY 66 ----- 65
Db 121 PPGASQTIKSTKRSPKPKKTKKVIKIESEITEHSHSVENQSSSSSSSSSTIR 180

```

```

QY 66 -----VKDNKKNRKTKKPPKPPVDEAGSGDNGDFKVTPTDST 106
DB 181 KIKSSKSAANRELQKLVKDKKKRRTKKKPPKPPVDEAGSGDNGDFKVTPTDST 240
QY 107 TOHNKYSTSKITTAAPINRPSLNPNSDSKSTSLTVNKETTVEKETTNNKQSTIDG 166
DB 241 TOHNKYSTSKITTAAPINRPSLNPNSDSKSTSLTVNKETTVEKETTNNKQSTIDG 300
QY 167 KEKTSAKETQSIEKTSANDLAPTSKVLAKPPKKAETTKGSAITLTPKRPPTTTPKEPAS 226
DB 301 KEKTSAKETQSIEKTSANDLAPTSKVLAKPPKKAETTKGSAITLTPKRPPTTTPKEPAS 360
QY 227 TTPKEPTPTTKSAPTTPEKAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEP 286
DB 361 TTPKEPTPTTKSAPTTPEKAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEP 420
QY 287 APTTTSAPTTPEKAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 346
DB 421 APTTTSAPTTPEKAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 480
QY 347 EPAPTAAPKAPATTPEKAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 406
DB 481 EPAPTAAPKAPATTPEKAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 540
QY 407 TTKSAPTTPEKESPTTTPKEPAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEP 466
DB 541 TTKSAPTTPEKESPTTTPKEPAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEP 600
QY 467 APTAPKAPATTPEKAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 526
DB 601 APTAPKAPATTPEKAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 660
QY 527 PEKAPATTTPKKAAPTTPEKAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEP 586
DB 661 PEKAPATTTPKKAAPTTPEKAPATTTPKKEPAPTTTKEPAPTTTKEPAPTTTKEP 720
QY 587 APTTPKKAAPKELAPTTTKEPSTTSOKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 646
DB 721 APTTPKKAAPKELAPTTTKEPSTTSOKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 780
QY 647 TAPTTLKAPATTTPKKAAPKELAPTTTKEPSTTSOKPAPTTTKEPAPTTTKEPAPTTTKEP 706
DB 781 TAPTTLKAPATTTPKKAAPKELAPTTTKEPSTTSOKPAPTTTKEPAPTTTKEPAPTTTKEP 840
QY 707 KPAATTPPEPTTSEVSTPTTKEPPTTHKSPESTPELSAETTPRALENSPEKPEVPT 766
DB 841 KPAATTPPEPTTSEVSTPTTKEPPTTHKSPESTPELSAETTPRALENSPEKPEVPT 900
QY 767 TKTAPATKPEKTTTAKDKTERDLRTTPETTTAAPKMKTEATTTTEKTESKITATTTTOY 826
DB 901 TKTAPATKPEKTTTAKDKTERDLRTTPETTTAAPKMKTEATTTTEKTESKITATTTTOY 960
QY 827 TSTTODTTPKKTITLTKTTTLAPAVTTTKTITTTTEIMNPEETAKKDRATNSKATTPK 886
DB 961 TSTTODTTPKKTITLTKTTTLAPAVTTTKTITTTTEIMNPEETAKKDRATNSKATTPK 1020
QY 887 POKPTAPKPKPTSTKPKKTPAVRKPKPTTTPPKMTSTMBELNPTSRISIAEMQOTTRP 946
DB 1021 POKPTAPKPKPTSTKPKKTPAVRKPKPTTTPPKMTSTMBELNPTSRISIAEMQOTTRP 1080
QY 947 OTTPSKLVEVNPKESEAGGEGTTPHMLRPHVMEVPTPDMDYLRVNOGIIIMPMLS 1006
DB 1081 OTTPSKLVEVNPKESEAGGEGTTPHMLRPHVMEVPTPDMDYLRVNOGIIIMPMLS 1140
QY 1007 DETNICKGKRVDELTLRLNGTLVAFRGHYFWMLSPPSPSPARILTEVNGISPIDVT 1066
DB 1141 DETNICKGKRVDELTLRLNGTLVAFRGHYFWMLSPPSPSPARILTEVNGISPIDVT 1200
QY 1067 RCNCEGTEFFKFSQWRFTNDIKDAGYKPKPTFGGGLTGOIIVALASTAKYKNNPESEY 1126
DB 1201 RCNCEGTEFFKFSQWRFTNDIKDAGYKPKPTFGGGLTGOIIVALASTAKYKNNPESEY 1260
QY 1127 FFKRGSIOQYIYKQEPVQKCPGRPALNVPYVGEMTOVRRRRFERAIGSOTHTIRIOY 1186

```

```

DB 1261 FFKRGSIOQYIYKQEPVQKCPGRPALNVPYVGEMTOVRRRRFERAIGSOTHTIRIOY 1320
QY 1187 SPARATODKGVNLNKKVSTLTMGLPNVYTSALSLPNIRKPDGYDYAFSKQYNYIDV 1246
DB 1321 SPARATODKGVNLNKKVSTLTMGLPNVYTSALSLPNIRKPDGYDYAFSKQYNYIDV 1380
QY 1247 PSRTARATITRSQOTLSKWYNCP 1270
DB 1381 PSRTARATITRSQOTLSKWYNCP 1404

RESULT 3
QJUM99 PRELIMINARY; PRT; 1054 AA.
AC QJUM99;
ID 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE MKNA, COMPLETE CDS, SIMILAR TO MEGAKARYOCYTE STIMULATING FACTOR
DE PRECURSOR AND CARTILAGE SUPERFICIAL ZONE PROTEIN.
GN PRG4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Ikegawa S., Nakamura Y.;
RT "a novel mouse gene highly homologous to a human gene encoding
RT megakaryocyte stimulating factor precursor and cartilage superficial
RT zone protein."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB034730; BAA92310.1; -.
DR MGD: MGI:1891344; Prtg4.
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR002965; P_rich_extensn.
DR InterPro: IPR001212; Somatomedin_B.
DR Pfam: PF01033; Somatomedin_B.2.
DR PRINTS: PR01217; PRICHEXTENSN.
DR PRINTS: PR00022; SOMATOMEDINB.
DR SMART: SM00120; HX; 2.
DR SMART: SM00201; SO; 2.
DR PROSITE: PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE: PS00524; SOMATOMEDIN_B; 2.
SQ SEQUENCE 1054 AA; 115991 MW; 4fC64BFA42283235 CRC64;

Query Match 49.4%; Score 3369; DB 11; Length 1054;
Best Local Similarity 49.7%; Pred. No. 2e-218;
Matches 701; Conservative 54; Mismatches 158; Indels 498; Gaps 23;

QY 1 MAMKTLPIYLILLLSVFYIQOVSSQ----- 25
DB 1 MAMKTLPIYCLSLILFVYLIIQOVSSQDSSCAGRCGEYSRDATCNCDYNCQHYMECCPDF 60
QY 26 -----ELSCGRCPESFERGECDCDAQCKKYDKCCDYESFCAB----- 65
DB 61 KRVCSPELSCGRCPESFARRECDQSCQKQKCCADYDSFCEVHNSTSPSSKTAAPT 120
QY 66 ----- 65
DB 121 PAGASDTIKSTTKSPKSPPTTRTIKVVESSELTREHSDSENOESSSSSSSSSTIRIKS 180
QY 66 -----VKDNKKNRKTKKPPKPPVDEAGSGDNGDFKVTPTDST 109
DB 181 SKNSANRELQKLVKDKKKRRTKKKPPKPPVDEAGSGDNGDFKVTPTDST 240
QY 110 NKVSTSKITTAAPINRPSLNPNSDSKSTSLTVNKETTVEKETTNNKQSTIDG 169
DB 241 SKVATSPKTTAAKPVTPKPSLAPNSETSKESLASNKETTVEKETTNNKQSSA-SKKK 299
QY 170 TTSKAKETQSIEKTSANDLAPTSKVLAKPPKKAETTTGSAITLTPKKEPTTTPKEPAS 229

```


DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE Y51811A.1 PROTEIN.
 GN Y51811A.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NC NCB1_TaxID=6239;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None.
 RC STRAIN-BRISTOL N2;
 RL "Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Cotton M.;
 RT "The sequence of C. elegans cosmid Y51811A.";
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [13]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [14]
 RP EMBL; AC006797; AAF60743.1;
 DR InterPro; IPR002965; P_fich_extensn.
 DR PRINTS; PR01217; PRICHEXTENS.
 SQ SEQUENCE 1079 AA; 110532 MW; 8BDBE3824CF80CA1 CRC64;

Query Match 14.3%; Score 971.5; DB 5; Length 1079;
 Best Local Similarity 29.5%; Pred. No. 2.8e-57;
 Matches 330; Conservative 129; Mismatches 481; Indels 177; Gaps 44;
 65 EVKDKKNNRKKKKPPKPPVVDGSGLDNDDFY-----TPDDSTTO-HNKVSTSP 116
 DB 31 ELADCRAGQTHMTP-----SILTSYDMETPSTLVLSPTSSSPPIKETTTPAR 81
 QY 117 KITTKAPINPPSLPNSDTSKESLTVNKEET-VETKETTTPKQTSOKKETSKE 175
 DB 82 ETTSTEP-----PSSSTTPVQTTTAPETTSAPSSSTTPVQTT-----TTTAP 128
 QY 176 TOSIEKTSADKLAPTSYKLAFTPKAETT-TKGP-ALTTKEPTTPPKPEASTPREPT 233
 DB 129 TTSIEPPSSS-----TSPVQTTTTPAETTSIEAPSSSTTPVQTTTAPETTSIEPPSS 184
 QY 234 PTTIKSAPTPPKPEAPPTTKSAPTPPKPEAPPTTKPEAPPTTKPEAPPTTKS 293
 DB 185 TSPVQTTTTPAETTSIEPPSSSTTPVQTTTAPETTSIEPPSSSTTPVQTTTAP 240
 QY 294 APTPKPEAPPTTKKAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTAP 353
 DB 241 APETTSIEPPSSSTTPVQTTTAPETTSIEPPSSSTTPVQTTTAPETTSIEPPSS 299
 QY 354 KKPAPTPPKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKS 399
 DB 300 TTPVQTTTTPAETTSIEPPSSSTTPVQTTTAPETTSIEPPSSSTTPVQTTTAPET 359
 QY 400 TK--EAPPTTKSAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTK 456
 DB 360 TREPPSSSTTPVQNTTTPAETTSIEPPSSSTTPVQTTT-----APETTSIEPPSSST 415
 QY 457 EAPPTTKKAPAPAPKEP--APTTKEPAATTPKALPTTPPEKLAATTPPEKAPPTPEEL 514
 DB 416 TTPVQTTTTPAETTSIEPPSSSTTPVQTTTAPETTSIEPPSSSTTPVQTTTAP 471
 QY 515 APTTPEPTTPPEAPPTTKAAPTTPKEP--APTTKEPAATTPPKPEAPPTTKPEAPPT 572
 DB 472 PETTSIEPPSSSTTPVQTTTTPAETTSIEPPSSSTTPVQTTTAPETTSIEPPSS 529
 QY 573 TTPKGTAPITLKEPAATTPKPKAPKELAPTTTKEPTSTSDKP-APTPKGTAPPTPKPE 631

DB 530 TTPVQTTTTPAETTSIEPPSSSTTPVQTTTTPAETTSIEPPSSSTTPVQTTTAPET 588
 QY 632 APTPKPEP--APTPKGTAPITLKEPAATTPKPKAPKELAPPTTKGPTSTSDKP-APT 688
 DB 589 --TTSIEPPSSSTTPVQNTTTPAETTSIEPPSSSTTPVQTTTAPETTSIEPPSSST 646
 QY 689 PKETAPTPPKPEAPPTTKPEK-----PAPTPPEPTTP-----EVSPTTTPKEPTT 735
 DB 647 PVQTTTTPAETTSIEPPSSSTTPVQTTTTPAETTSIEPPSSSTTPVQTTTAPETTS 706
 QY 736 HKSDSTPELSAAPT--PALENSPEKPGVPTTKPAATKPEMTTAKKTERDTRTT 793
 DB 707 TEPPSSSTTPVQTTTTPAETTSIEPPSSSTTPVQTTTTPAETTSIEPPSSSTTPVQTT 766
 QY 794 ----PETTAPAPKMTKEATTTTEKTSKITATTTQVTSITTTODTPPKITTKTTLAP 849
 DB 767 TTPAETTSIEPPSSSTTPVQTTTTPAETTSIEPPSSSTTPVQTT--TTAPETTSIEP 824
 QY 850 KVTTKTKTITTTTEIMNKPEETA-KPKDRATNSKATTPPKPKPKAPKPKSTKKPTMPR 908
 DB 825 PSSSNTPEVQTTT--TTAPETTSIEPPSSSTTPVQTTTTPAETTSIEPPSSSTTPVQTT 882
 QY 909 VRKPKTPPRKMTSTMPELNPTSRILAEMLQTTTPRPNQNSKLVNPKSEDAGAGC 968
 DB 883 ITAPETTSIEPPSSSTTPVQTTTTPAET--TSTEP--PSSSTTPVQTTTTPAETTS 936
 QY 969 -ETPHMLRPHVMPETP---DMDYLPVY-----PNCGLIINMLSDEN----- 1010
 DB 937 TEPPSSSTTPVQTTTTPVDCSLSTIDRVYFTTEEMENKRDITIOSYDSPRTAFSE 996
 QY 1011 -----ICNGKPYDGLTTLRNGTLVAFRGHYFMLS-----PSPSPPAR 1050
 DB 997 VSFEDIGCTAILTCITYSEGGISNL-NMTL-----FGLSDSGSIDLP----- 1040
 QY 1051 IFEWVGISPIDVTFRCNCEGKTFEFDKSDQYWRFTN 1087
 DB 1041 ----YVNPGLG-IMPEINCEGK-----WSTIN 1064
 RESULT 6
 ID 0917S1 PRELIMINARY; PRT: 1049 AA.
 AC 0917S1;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE CG5228 PROTEIN.
 GN CG5228.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCB1_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D.; Celniker S.E.; Holt R.A.; Evans C.A.; Gocayne J.D.;
 RA Amanatides P.G.; Scherer S.E.; Li P.W.; Hoskins R.A.; Galle R.F.;
 RA Gnanapavan R.; Lewis S.E.; Richards S.; Ashburner M.; Henderson S.N.;
 RA Sutton G.G.; Wortman J.R.; Yandell M.D.; Zhang D.; Chen L.X.;
 RA Brandon R.C.; Rogers Y.-H.C.; Blaise R.G.; Champe M.; Pfeiffer B.D.;
 RA Wan K.H.; Doyle C.; Baxter E.G.; Helt G.; Nelson C.R.; Miklos G.L.G.;
 RA Abell J.F.; Agbayani A.; An H.-U.; Andrews-Pranichoch C.; Baldwin D.;
 RA Baller R.M.; Basu A.; Baxendale J.; Bayraktaroglu L.; Beasley E.M.;
 RA Beeson K.T.; Benos P.V.; Berman B.P.; Bhandari D.; Bolshakov S.;
 RA Botkov D.; Botchan M.R.; Bouck J.; Brocksen P.; Brothier P.;
 RA Butty K.C.; Busam D.A.; Butler H.; Cadieu L.; Center A.; Chandra I.;
 RA Chertys J.M.; Cawley S.; Dahlke C.; Davenport L.B.; Davies P.;
 RA de Pablos B.; Delcher A.; Deng Z.; Mays A.D.; Dew I.; Dietz S.M.;
 RA Dodson K.; Doup L.E.; Downes M.; Dugan-Rocha S.; Dunkov B.C.; Dunn P.;
 RA Durbin K.J.; Evangelista C.C.; Ferraz C.; Fertler S.; Fleischmann W.

RA Foster C., Gabriellan A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Malshina N.V., Moharry C., Morris J., Mosheret A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RA Science 287:2185-2195(2000).
 RA EMBL: AE003495; AAG22353.1;
 RA FlyBase: FBgn0030561; CG5228.
 DR SEQUENCE 1049 AA; 107278 MW; 954DD629E7619671 CRC64;

Query Match 13.7%; Score 936.5; DB 5; Length 1049;
 Best Local Similarity 34.6%; Pred. No 6.1e-55;
 Matches 371; Conservative 45; Mismatches 439; Indels 217; Gaps 54;

OY 67 KDNKRNRTKKRPKPPVVDAGSGDNGDFKVTTPDTS---TQOHNVSTSPKITTAKP 123
 DB 30 KNEHGTAKPTTLKP-----TEGSAKPTTLKPTEGTSAKPTTLKP 71
 OY 124 -----INPRSPSPNSDSKESL-----YNNKETTVEKETTNNKQT--STGCKEKT 171
 DB 72 TEGTTAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGTTAKPTTLKPTEG 127
 OY 172 SAKET--GSIEKTSKAK--DLAPTSKVLAKPT-----PKAETTCKGALTPKEPPTT 220
 DB 128 SAKPTTLKPTEGTTAKPTTLKPTEGTSAKPTTLKPTEGSAKPTTLKPTEGTSAKPTTLK 187
 OY 221 PRE-----PASTPKEPT---PTTKSAPTTKEPAPTTTSAPTTKEPAPTTKEPA 272
 DB 188 PTEGTSKAPPTTLKPTEGTSKAPPTTLK--PTEGTTAKPTTLK--PTEGTSKAPPTTLKPTEG 243
 OY 273 TTPKEPAPTTKEPAPTTTSAPTTKE--PAPTTPKPAPTTKEPAPTTKEPT---P 327
 DB 244 TTKA---PTTLKPTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTTAKPTTLKPTEGTSAKP 300
 OY 328 TTPKEPAPTTKEPAPTTKEPAPTLAKKAPPTTKEPAPTTKE-----EPAPTTPKEPSP 382
 DB 301 TTKLPTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTTAKPTTLKPTEGTSKAPPTTLKPTEG 360
 OY 383 TTPKEPAPTT-----TKSAPTTTKEPAPTTAKSA--PTTKEPSPPTTKEPAPTTKE 433
 DB 361 TTKA---PTTLKPTEGTSKAPPTTLKPTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTTAK 416
 OY 434 PAPTTPKPAPTTTPKEPAPTTKEPAPTTTPKKA--PAPKEPAPTTTPKEAPTT--PK 488
 DB 417 --PTTLKPTEGTTAK--PTTLK--PTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTTAKPT 470
 OY 489 KLAFT-----TPEKLAFTTPEKAPPTTPELAFT--TPEEPPTTPEEPAPPTPKAAAPN 541
 DB 471 TTKLPTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEG 530
 OY 542 TTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKKA--PTTKEPAPTTTP 591
 DB 531 KGTAKPTTLK--PTTEGTSKAPPTTLKPTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTTA 587
 OY 592 K---KPA-----PKELAPT-----TTKEPTSTTSOKPA-----PTTPKGT 623

DB 568 KPTTLKPTEGTSKAPPTTLKPTEGTSKAPPTTLKPTEGTTAKPTTLKPTEGTSKAPPTTLKP 647
 OY 624 APTTKEPAPTTTPKPAPTTTPKKAPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTTAK 679
 DB 648 EETTKA---PTTLK--PTEGTSKAPPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTTAK 701
 OY 680 TSDKAPPTTPKE--TAPTTPKEPAPTTTPKKA--PTTEPPTTTPKEPPTTTPKEPPTTTPKE 733
 DB 702 TSKAPPTTLKPTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKP 761
 OY 734 TTKSPDSTPELSAAPT-----KALENSPKEPV--PTYT---PAATK 774
 DB 762 TL--KTEBT---SAKPTTLKPTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTSKAPPTTLK 816
 OY 775 PEKTTAKOKTERDLRT-----EETTAAKMKREKAT-----TEKTTES 817
 DB 817 PTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTTAK 876
 OY 818 KITATTTQVSTTTQOTTPFKITTLKTTTLAKPVTTTK--TITTEIMNKPEETAKPKD 875
 DB 877 PTLKPTEGTSKAPPTTLKPTEGTTAKPTTLKPTEGTTAKPTTLKPTEGTSKAPPTTLKPTE 936
 OY 876 RATNSKATTPKPKPKPTKAK--KPT--STKKPKTPPVRAKKTPT--PRKMTSTMP- 926
 DB 937 -GTAKPTTLKPTEGTSKAPPTTLKPTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTSKAPPT 995
 OY 927 ELNPSTRIAEAMQTTTPRNPOTNSKLVENPKSEDAGAGC-ETPHMLLRP 977
 DB 996 TLKPTERTSAQ--PTTLKPTEGTTAKPTTLKPTEGTSKAPPTTLKPTEGTSKAPPTTLKP 1045

RESULT 7
 ID 076894 PRELIMINARY; PRT: 1795 AA.
 AC 076894;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE EG:5667.1 PROTEIN.
 GN EG:5667.1 OR CG14796.
 OS *Drosophila melanogaster* (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY.
 RA MEDLINE-20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Abuli J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley S.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotler P.,
 RA Bokkova D., Botchan M.R., Bouck J., Broksstein P., Brotler P.,
 RA Burris K.C., Busam D.A., Butler G., Cadien E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davison L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dudin K.J., Evangelista C.C., Ferraz C., Ferris W.M., Glasser K.,
 RA Foster C., Gabriellan A.E., Garg N.S., Guan P., Harris M.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

QY 186 DLAFSTVLAKPPTKPAETTTKGPALTPKREPTTPKPEASTTPKREPTTTIKSAPLTPK 245
 DB 478 TYAPTKETTYAP---EETTYASTEETTYAPTEETTYAPAEETPYEPTTEET-TYAPTEET 533
 QY 246 EPAPT-TTKSAPPTPKREAPPT--TTKEPAPPTKREAPPTTKREAPPTTKSAPT----- 296
 DB 534 TYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETT-YAPTEETMYA 592
 QY 297 ---TPKEPAPPTPKKAPPTPKREAPPTTKREPTTPKRP-----AP--- 335
 DB 593 PTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYAPTE 652
 QY 336 TTKEPAPPTPKREP-----APTAKKAPPTPKREAPPTTKREAPPT--TTKESPTTPKE 387
 DB 653 TYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYE 712
 QY 388 P-----APT-TTKSAPT-----TKEP-----APT-TTKSAPT--TKESPT 421
 DB 713 PTEETTYAPTEETTYAPAEETMYAPTEETTYGPTTEETTYAPTEATTYAPTEETTYAPTE 772
 QY 422 TTKEP-----APTPKREAPPTTKKAPPTPKREAPPT--TPKREAPPTTKKAPPTAPKE 473
 DB 773 TTYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYE 832
 QY 474 PAPTPKREAPPTPKKLPPTPEKLAPTTPKREAPPTPEELAPT--TPPEPTTPTPPEP- 530
 DB 833 PTEET--TYTPTEETTYAPTEETTYAPTEKTTTYAPTEETTYAPTEETTYAPTEETTYAP 890
 QY 531 -----APT-----TPKAAANTPKREAPPTPKRE-----APTPKREAPPTPKRE 569
 DB 891 KETTYAPTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEET 950
 QY 570 TAPTPKGAAPT--TLKEPAPPTPKKAPK-ELAPT--TTKEPTSTSDKAPPTPKCA 624
 DB 951 YAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1010
 QY 625 PTPPKREAPPTPKRE-----APTTPKGAAPT-----TLKEPAPPTP 660
 DB 1011 EETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYASTEETTYAPTEETTYAPAEETP 1070
 QY 661 KRP-----APK-----LAPT--TTKGPSTSTSDKAPPTPKETAPTTPKREAPPTPKKPA 709
 DB 1071 YETTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAP 1130
 QY 710 PTPPEPTTPPTSVESTPTTKPT--TIKSPDESPTPELSAPTEPKA-LENSPKERG-- 763
 DB 1131 EETTYAPTEETMYAPTEETTYGPTTEETTY-APTEATTYAPTEETTYAPTEETTYAPTE 1189
 QY 764 --VPTKTPAPATKPEMTTAKDKTERDLRTP---ETTTAPKMTKETATTTKETS 817
 DB 1190 TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1244
 QY 818 KIRATTTQVSTTTQDT-----PKKIT--LKTTLAPKVTTKTITTELM 864
 DB 1245 TYAPTEETTYAPTEETTYAPTEETMYAPTEETTYGPTTEETTYAPTEATTYAPTEETTPA 1304
 QY 865 NKPEETAKKORDAPT--NSKATPKP-OKPTKAPKPT-----STKRP-KTMR 908
 DB 1305 PTEETTYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1364
 QY 909 VRKKTPTPKMTSTMPBELNPTSRIAEAMLOTTTP--NQTNSKLVEYNKSEAG 964
 DB 1365 TDEPTDEPTDE--PSDEPTDEPTDEPTDLPTDEPTSPCDNOGINGIGVENKRYNAG 1420

RESULT 9
 Q9SPMO PRELIMINARY; PRT; 1315 AA.
 AC Q9SPMO;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE EXTENSIN-LIKE PROTEIN.

GN PEK2.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 OC NCBI_Taxid=4577;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=POLLEN.
 RA Stratford S., Barnes W., Golubiewski A., Colter R., McCormick S.,
 RA Horst D., Gao M., Showalter A., Bedinger P.A.;
 RT "Pollen Extensin-Like (Pex) Genes in a Monocot and a Dicot."
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF159297; AAD5980.1;
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR001998; Xylose_isom.
 DR InterPro: IPR002965; P-rich_extensin.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR; 3.
 DR PRINTS: PR01217; PRICEXTENS.
 DR PROSITE: PS00172; XYLOSE_ISOMERASE_1; UNKNOWN_1.
 DR SMART: SM00370; LRR; 3; 134401 MW; 64C97A2A01F0936F CRC64;
 SO SEQUENCE 1315 AA; 134401 MW; 64C97A2A01F0936F CRC64;

Query Match 11.9%; Score 811.5; DB 10; Length 1315;
 Best Local Similarity 24.6%; Pred. No. 2e-46;
 Matches 262; Conservative 89; Mismatches 473; Indels 243; Gaps 35;

QY 32 RCESEERGRCCDCKKIDKCCPDYSEFCAEVNDKNNKTKKKPT-----K 81
 DB 390 QCAPVLARPEVDCSKH-----VCAGY-----PTGGGPPSSSVRK 425
 QY 82 PPAVDEAGSLDNGDFKVTTPDSTTQHKNVSTPKITTKAPINRPSLP----- 132
 DB 426 PPSVPG-----KRAAPAMPPTPHRPDVS-----EPL-PPSPVPAPAPWRMT 469
 QY 133 -NSDTSKETSILVKNKETTVEETKETTINKOTSTDGKETSNAKEIOSIKTSKDLAPTS 191
 DB 470 LRSPADEIYPT--PPVPAKSPPGTSPASGAPPLAQAPPAASSPATIPYKSSPPA 525
 QY 192 KV-----LAKPTKAEETTKGPAITTEKEPTTPPKREAPPTPKREPTPTI 237
 DB 526 AVVLEPPAKPTSPAPAPASPPPEAPVSSPQOVKSPPPAPAVASPPPMKSP--PPAPV 583
 QY 238 KSAPTPKREAPPTTKSAPTP-KEAP-----TTKREAPPTTKREAPPTTKREAPPT 289
 DB 584 ASPPPLMKSPPPAPVASSPQPLKSPPPVLMSTPSKSPPPVAVASPPPVKSPPL 643
 QY 290 TTKSAPTPPK-----PAP--TTP-----KRPAPTPKREAPPTPKRE--PTPTPKREAPPTK 338
 DB 644 APVSSPPSVKLPPLPAPAGKSTPPPEEKPTPTTPVKSPPPEKSLPPTLTSSPPOEK 703
 QY 339 EPAPTPKREAPPTAKKREAPPTPKREAPPTPKREAPPTTKESPTTKREAPPTTKSAPT 398
 DB 704 PTPETPKPPPPSPVETLPPKSSPPEEVSPPAPAKSSSP-----PAPVS--SP 756
 QY 399 TTKEPAPPTTSAPPTPKREPSPTTKREAPPTPKREAPPTPKKAPPTPKREAPPTPKRE 458
 DB 757 LKSSPPPVESPPTPKSSPPLAPVSSPOVEKTSPPAPVSSPPTPKSSPPLAVSS 816
 QY 459 APTTKKAPPTAPKREAPPTTKREAPPTPKRLPTTPPEKLAPTTPKEPAPPTPEELAPT 518
 DB 817 PPOVEKTSPPAPVSSPPTPKSSPPLAPVSSPOVEKTSPPAPVSSPPLPKSPSPS 876
 QY 519 PEEPTTPPEEPAPPTPKAAAPNTPKREAPPTPKREAPPTPKREAPPTPKREAPPTPKGT 578
 DB 877 SVSSPPTTVKSSPPAPLSSPMTKSSPPAHVSSPEAKSSPPLATISSPSEKSP 936
 QY 579 A-----PTTKREAPPTPKKPAKELAPTTKEPTSTSDKAPPT-----TPKG 622
 DB 937 SSPPMVEKTSPPAPVSSPPTPKSSPP-----APVSSPPVAVKSSPPAPVSSPPTPKP 993

QY 623 TAPTP-----KEBAPTTPKEBAPTTPK--GTAPTTKEBAPTTPKKAPKELAPT 672
 Db LPPAPVSPPPVKKSSPPPTPVSSPPPTPKPLPPPTPVSSPPPTXKPLPPP--APVS 1050
 QY 673 TKGPTSTSDKRAPPT--PKETAPTPK-----EPAPTTPKRAPPTPEPTTSEVSTP 726
 Db 1051 SPPEVKKSSPPAPVSLPPPTKRPSPPTRVSSPRPVKCCPPPTLVSSPPAPKSLPPP 1110
 QY 727 TTTKEPTTIHKSPESTPELSAPPTPKALENSPKKEGCVPTTKTPATKEBMTTAAKDTT 786
 Db 1111 TPVSSP-----PPEVKKSSPPPTPVSSPPAPKSSPPPTPVSSPP----- 1150
 QY 787 ERLKRTT-----ETTAAPKMTKETATTEKTESKITATTOVSTTODTPPKITT 841
 Db 1151 --ELKSSPPAPVSSPPAPKSSPPAPVSLPPEVKKSSPPAPLSSPPAPKSP----- 1203
 QY 842 LKTTTAPKVTYTTKTTTTEINMKKEEPAKRDATNSKATTPKOKATKAPKKTSTK 901
 Db 1204 -----PPRA-----PSSLPPEVKKSSPPAPVSSPPAPVSSPPAPKSP 1242
 QY 902 KPKTPMRVKRKTTPPRKMTSTMPELNPTSRIAEAMLTTPRNOTPNSKLVENPKSE 961
 Db 1243 SPPEAP--VKPSLPPAPVSSPPAPV-----TSAP-----PKKE 1275
 QY 962 DAGGAGCTPRHMLRHYVMEVTPDMVLPVNOGIIINPMSLDE 1008
 Db 1276 EDSA-----PPAELPPSPFNIDILPPIAMNK 1303

RESULT 10
 Q20007 PRELIMINARY: PRT: 1274 AA.
 ID Q20007:
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE COSMID F35A5.
 GN F35A5.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 GA Jones M., Kershaw J., Kirsten J., Laister N., Laticelle P.,
 LA Lingham J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden K., Sulston J.,
 RA Thierly-Wieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans".
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Leimbach D.;
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U46675; AAB52641.1; -
 SQ SEQUENCE 1274 AA; 138065 MW; 07F6BD0292C1799F CRC64;

Query Match
 Best Local Similarity 28.0%; Score 802; DB 5; Length 1274;
 Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

QY 71 KNRTKK--NP-----TP-----KPVVDEAGSGLNGD-----FKVTPDSTTGHNKVST 115
 Db 274 KNPTKKKKPWEDETPVEEVEKEPPVEKKAPVLKKBPAPAKARDSPSKAPKVEPS 333
 QY 116 PKITTAPEI--NPRSLPNSDTSKETSILTVKERTVEKETTNNKOSTSDGKEKTSK 174
 Db 334 SPVVPPTPVKNPVKKYKRPWEVDEDEPAEVEVKKBPAPKPTVLKRRKEPSSSTPSSDPS 393
 QY 175 ETOSIEKTSAKDLAPTSKVLAKPTPKAE-----TTTKGPA-----LTPP 213
 Db 394 PKAAPAVKPRDSSPKATPLADBPKAQEVPPPVKNPVKKYKRPWEVDEDEDEVEEVKQP 453
 QY 214 KEPTPTTP--KEPASTTPEK--TP--TTIKSAPTTKEBAPTTPKAPTTKEBAP 264
 Db 454 EAPAKKTPTVLKRRKPAKADAKPATSKTPTPEKDDVKKPRDSSPKVAAKPPSAAPA 512
 QY 265 TTTKEPA-----PTTKBAPTTTKEBAPT-----TTKSAPTPP 298
 Db 513 TPVKNPVKKMRPWEDETPADDSKPTDAKKTTPSLAKKDPAPAKESLKPADTKAPAP 572
 QY 299 KEP-----APTTPKKRAPTPKE-----RAPTTKEPTTPTPKBPAPTTKEBAPTTPK 348
 Db 573 RDSPPKVAAPTAPEKTPVLAKKEBAPADSKTEKEKSKRPSBKKAVPAPVKTVEV 632
 QY 349 APTAPKKRPAD-----TTKBPAPTTTKEBAPTTPKEP-----SPTPKBPAPTTKSAPT 398
 Db 633 APAVAKKPEPISKPDTPAKPAKKAEPNSVPV--PIPVKNPVKKMRPWEDDAPKAPVSLDE 691
 QY 399 TTKBPAPTTKSAPTTP--KEPSPTTKEBAPTTPK--EPAPTTPKKRPAD--TTKBPAPT 453
 Db 692 PEKK--TPVLAKKAPTPKPDSEAAADPVSGSPSKDPLAKKAPKPRDPSPMKAVPIKPAK 750
 QY 454 TPKEBAPTTPKKRAPTA--KEBAPTTPKETAPTTPKLTPTP-----EKLAPTPPEKPA 507
 Db 751 T--EVPAPVAKKPEPAKARDSPSKAK--AEPNSP--VPEPTPVKNPVKKMRPWEDDA 805
 QY 508 PTTPEELAPTTPEP-----PTTPEBAPPTPKAAAPNTPKBAPPTTPKEBAPTTP--PKE 561
 Db 806 PABPVAVPEPEKTPVLAKKTTPKPRDPSPKRAVPAKPESTKDAPVSKKREPVSKEKE 865
 QY 562 PAPTTPKETAPTPKGTATTTKEBAPTTPKKAPKELAPT--TTKEPTSTSDKAPPTTP 620
 Db 866 PSFKKAKRPSVVP-----TPVKNPVKKY--KRPWEDEDEPEEVEVKKSE--PEKTPVLA 918
 QY 621 KGTAPTTPKEBAPTTPKEBAPTTPKGTATTTKEBAPT-----TPKKRPAP-----KEL 668
 Db 919 K--KEPEKAPD--APKVAAKRDPSPKKAVE--KEBAPVAAKRPDLSKALPIRANQEA 974
 QY 669 APTTTKGPTS-----TSDKP-----APTTPKEI-----APTTPKBPAPTTPKKAPPT 711
 Db 975 PPTPVKNPVKKMRPWEDEDEPAEVSAPBEPKPTVLAKKAPKPRDP--SPKKAPV 1031
 QY 712 TPETPPTTSEVSTPTTKEPTTIHKS-----DESTEPISA--EP--TPKALENSPKPG 763
 Db 1032 AAK--DPAKPIEV--PPTPVKNPVKKMRPWEDEDESEPSVSAPEPEKTPVLAKKAPKPA 1089
 QY 764 V-----PTTKPAAT-----KEBMTTAAKDTTELDLKTTPETT--TAAPK 802
 Db 1090 TKPDEAAADVSGPTSDKPTLSKKAPEVKRPPTDPKDDKLSPAKKKEKAPAEPAAPK 1149
 QY 803 MKETATTTETTESKITATTOVSTTODTPFKITTLKTLAKRVTTTKTITTTT 862
 Db 1150 KWKPWDDDDPDEADFTVPAPSKKPPTEBAPDLG-----GPKTDPK----- 1193

```

QY 863 IMKPEETAPKDRATNSKATTPKPOKPTKAKPKPTSTKKTKMTPRYKPK-----913
DB 1194 -LNKKAPAEKPTKE-----PKPKVSKPPKPTKEPTPKP-AAFKMKRPMDDEDEPE 1243
QY 914 ---TTPPKMTSTMP 926
DB 1244 ADFTMPAKPKDTEDP 1259

RESULT 11
09LIE8 PRELIMINARY; PRT; 1480 AA.
ID 09LIE8
AC 09LIE8
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE SIMILARITY TO CELL WALL-PLASMA MEMBRANE LINKER PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety PL,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL; AP001306; BAB03062.1;
DR InterPro; IPR002965; P_fich_extensn.
DR PRINTS; PRO1217; PRICHEXTENS.
SQ SEQUENCE 1480 AA; 147153 MW; DIA0C079F155E732 CRC64;

Query Match 11.7%; Score 800.5; DB 10; Length 1480;
Best Local Similarity 29.6%; Pred. No. 1.2e-45;
Matches 307; Conservative 54; Mismatches 505; Indels 171; Gaps 45;

QY 75 KKKRTPKPVNDEAGSLDNGDFKVTTPDSTQHNKSTVS---PKTTAKPIMRPSL 130
DB 117 KRRPHKRPVVK-----RPHNPKRPLKRNHNPKRPTKNNHNPKRPTIKRPPKPSV 167
QY 131 PPNSDTSKETSGLTVNKETVEKETTNTKOSTDCKEKTSAKESQISEKTSADLAPT 190
DB 168 KRPSTPK-----RPPTNPBSTPQRPNNKRPSTPTPVASPPKAPT 211
QY 191 SKY--LAKPKAEKTTKGPALTTPKEKP--TTP--KEPASTTKEPPTTIKSAPTP 244
DB 212 TQMPPIATP-PIAKSPVATPPIATPPIATPPIIPVATPPIITPPIAMPPIATP 270
QY 245 KEPAPTT-----KSAPTPKEPAPTT-----KEPAPTPKEPAPTTKE 285
DB 271 PVAPAPITNPISKPVTTP-----PTTTPPIAKPPIATPPISTPPATPPAATPPIITP 326
QY 286 PA-----PTTKSAPTTP-KEPAPTTPKKPAP--TTPKEPAPTTPKEPAPTTPKEPAP 335
DB 327 PAKPVAISPIVT--PPTVIAQPVATPPTAPPVATPPIATPPIKSPISSTISPTPSES 384
QY 336 TTKPAPPTP-KEPAPTAPKPPAPTTPKEPAPTTPKEPAPTTP-----KEPSTTPKEPAP 390
DB 385 VAPPTATSPKTPPKAPKPPVATPPIAKSIATPPTATPPVATPPILEKPPVATPPTTP 444
QY 391 TTKSAPTTTKEPAPTTSAPTTPKEP--SPTTKPAPPTTPKEPAP--TTPKPAPT 445
DB 445 TAT---PVAKPPVETPPIATPPTPAKPPISSTPISKPPVATPPAATPPIITPPIVAPVA 501

```

```

QY 446 TKEPAPTTPKEP--APPTTKKPAPTAP--KEPAPTTPKETAPT-----PKKLP--T 493
DB 502 TPLAIPVAKPPVVTPTATPPIATPPIAKSPVATPPTATPPVATPPIAKPPVITPTT 561
QY 494 TPEKLAPTTPKEPAPTTPBELAPT--PEEPPTTPEEPAPTTPKAAAPT--PREAPT 549
DB 562 TPLATPVPKPPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPI 621
QY 550 TKEPAPTTPKEPAPTTPKETAP--TTPKGTAPT--LKEPAPTTPKPAKELAPT 604
DB 622 TPLATPPIAKPPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPI 681
QY 605 KEPTSTSKPAPPTTPKGPAP--TTPKEPAPTTPKEPAPTTPKETAPTTPKEPAPTTPK 663
DB 682 KPPVAT--PPTATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIAT 736
QY 664 APKELAPTTTKGPTST--TSDKPAPTTPKETAPT--TKEPAPTTPKKAAPTTP-----E 714
DB 737 MPPIATPPIAKPPVATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPI 796
QY 715 TTPPTTSEVSTPTTKEPTTIH--KSPDESTBELSAEPTPKALENSPKKPGVPTTKTPAA 772
DB 797 TPIAKPPVATPPTATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIAT 856
QY 773 TKPEMT--TKADKTEBDLTTPETTTAAPKMTKETATTTTEKTSKITATTOVSTTT 831
DB 857 TPIITTPPKAPPVATPPIATPPI--IAKPPVATPPIATPPIATPPIATPPIATPPIATPPI 914
QY 832 QDTPPEKITTLLKTTIAFKVTTTKT---ITTEIMNKPDEETAKPKDATSKATTPK 887
DB 915 PAKPVAIPPIATPPIATPPIAKPPVATPPIATPPIATPPIATPPIATPPIATPPIATPPI 971
QY 888 QKPKAKKPKPTSTKKTPMRYKPKPTTPPKMTSTMBELN-----TSRLAE--- 936
DB 972 KPTTTP--PLAT--PPIVAMPPIATP--PIAKPPVATPPIATPPIANPVEKPPVATPPIAP 1025
QY 937 ----AMIQTTTRPNQTPNSKL--VEVNPESDAGAGETPMLLRPHVMEVTPDM 989
DB 1026 VLPPIAKPPVETSPATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPIATPPI 1070
QY 990 YLPRVNPQGIINPMLS 1006
DB 1071 ATPPTNPPIAMPPIVT 1087

RESULT 12
ID 041805 PRELIMINARY; PRT; 1188 AA.
AC 041805;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE EXTENSIN-LIKE PROTEIN PRECURSOR.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BJ7; TISSUE=POLLEN;
RA Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z34465; CA84230.1;
DR Mendel; 14346; Zeama; 2368; 14346.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003592; LRR_out.
DR Pfam; PF00560; LRR; 3.
DR SMART; SM00370; LRR; 4.
DR SIGNAL.
KW SIGNAL.
FT SEQUENCE 1188 AA; 120980 MW; 2C77C7F8D7130149 CRC64;

```

Query Match 11.5%; Score 784; DB 10; Length 1188;
Best Local Similarity 29.9%; Pred. No. 1.2e-44;
Matches 230; Conservative 59; Mismatches 359; Indels 122; Gaps 33;

QY 78 PTPKPPVNDAGSGLDNGDFKVTTPSTTQHNKVSSTPKITTAAPINRPSL-PPN 133
DB 460 PPHSSPPAD-----DYVPTPPVPGKSPATSPSPQVPPAASPPSLVKLSPPQ 510
QY 134 SDTSKETSILVNKETVETKETTITNKQSTDSKETSKEQSTIEKTSAKLAPTSKY 153
DB 511 APVG-----SPPPVKTTSPADIG---SPSPPPSVV 541
QY 194 -----LAKPTPKAETTKGPAITTKPEPTTKPEASTPKPEPTTKISAPTTKEPA 248
DB 542 SPPPVKSPPPAPVSGSPPEKSPPPAPVAVASPPPVKSP-PPPTLVASPPPVKSP 599
QY 249 PTTTKSAPTP--KEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTP--KEAPTT 305
DB 600 PPAAPVASSPPPVKSPPPPVASPPPPAPVASSPPPMKSPPTPVSSPPPEKSPPPP 659
QY 306 PKAPATTKPEAPTT-----PKE---PTPTPKPEAPTTKEAPTTKEAPTTAKKP 356
DB 660 PAKSTPPEEYPTPTSVKSSPPEKSLPPTLIPSPPEKPTPTSPKRP-SSSEKP 718
QY 357 APTPKPEAPTTKEAPTTKEAPTTKEAPTTKEAPTTKEAPTTKEAPTTKEAPTTK 416
DB 719 SP--PKEPVSSPPQTP-----KSSPPAPVSPPTPVSSPPALAPVSSPSVSSP 768
QY 417 EPSPTTKEAPTTKEAPTTKEAPTTKEAPTTKEAPTTKEAPTTKEAPTTKEAP 476
DB 769 PPAAPLSSPPAPQVKSPPVQVSSP--PPAKSSPLAPVSSPPQVEKTSPPAPLSSPP 827
QY 477 TTPKEAP-----TTP-----KKLTTPTEKLAAPTTEKAP-----TTPBELAAPTTEPT 523
DB 828 LAKSSPPHVVVSSPPPVKSSPPAPVSSPPAPVSSPPAPVSSPEVVKSTP--PA 885
QY 524 PTPPEAPATTPKAAPTKE--PAPTTPKEAPTTKEAPTTKEAPTTKEAPTTKEAPTT 582
DB 886 PTTVISP--PSEKSPSPPTPVSLPPIVKSPPAPVSSP--PMTPKSSPPPVVSSPPT 943
QY 583 LKE---PA-----PTTPKK---PAKELAPTTKE--PTSTSDKAPATTPKGATPTP 628
DB 944 VKSSPPAPVSSPPATPKSSPPAPVNLPPPEVKSPPPPVSSPPA---PKSSPPAP 1000
QY 629 -KSPATTPKEAPATTPKGATPTTKPEAPTTKEAPTTKEAPTTKEAPTTKEAPTT 687
DB 1001 MSSPPPEKSPPPAPVSSPPPVKSSPPAPVSSP--PPVKSPPPPAPVSSPPPV 1057
QY 688 -TPKETAPTTPKPEAPTTTPKAPATTPEPTTSEVSTPTTKEPTTIKSSPESTPEL 746
DB 1058 KSPPPAPVSSPPPVKSSPPAPVSSP--SPPPVKSSPPAPVSSPPPVKSSPPAPV 1115
QY 747 SAEPPTKALENSPKPEGVPTTKTAAATKPEMTTAAKOTTERDLKTTPTT 796
DB 1116 SPPAP-----VKPPSLP--PPAPVSSPPVVPAPPKKEQSLPPAES 1158

RESULT 13
P70670 PRELIMINARY; PRT; 2187 AA.

AC P70670;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NASCENT POLYPEPTIDE-ASSOCIATED COMPLEX ALPHA POLYPEPTIDE (ALPHA-NAC,
NACA.
GN NACA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=6312450; PubMed=8698236;
RA Kotov W.V., St-Arnaud R.;
RT "Differential splicing in of a proline-rich exon converts alphaNAC
RT into a muscle-specific transcription factor.";
RL Genes Dev. 10:1763-1772(1996).
DR EMBL; U48364; AAB18734.1; -;
DR EMBL; U48363; AAB18732.1; -;
DR MGD; MGI:106095; Naca.
DR InterPro; IPR002715; NAC.
DR InterPro; IPR003037; TS-N.
DR Pfam; PF01849; NAC; 1.
DR Pfam; PF02094; TS-N; 1.
SQ SEQUENCE 2187 AA; 220599 MW; 003646AAB64DEBFD CRC64;

Query Match 11.4%; Score 776.5; DB 11; Length 2187;
Best Local Similarity 27.6%; Pred. No. 7.5e-44;
Matches 290; Conservative 119; Mismatches 416; Indels 227; Gaps 48;

QY 57 PDYESGCAEVKDKKKKNTKKKTPKPPVNDAGSGLDNGDFKVTTPSTTQHNK----- 111
DB 809 PKVDPLMSDVTPTSPKKTSAVAPK-----DTSATLSLKSVPAY 847
QY 112 VSTSPKITTAAPINRPSLPPNSDTSKETSL--TVNKET-----TVETKETTITNKOTS 163
DB 848 TSLSPKAPAPVNSENATIVTEIPTSILKNLAATPKETLATISPKYTSRSPKTSV 907
QY 164 TDGKEKTSKETSQSTIEKTSAKDLAPTSKVLAPTPKAETTTKGPAITTKPEPTTPKE 223
DB 908 LKGAAPMTSKKAT--EIAASKDVSPSQ--FPEVPLDQHV---PPSPSPSPVSDTLSC 959
QY 224 PASTTPEPTTITKSAPTPKPE--PAPTTKSAPTPK-----EPAPTTKEAPTT 273
DB 960 ALTSPPKGPATLAETPTPKSPKPAASKTTPATPSGVAVAVLEIPCSKAKATA 1019
QY 274 TPKEAPTTTKEAPTT--TKSAPT-----TPKEAPTTTTPK-P 309
DB 1020 APRESSATSSSKRAPKAVASKEIIPSKGVAVLEISLPEKETSATPGEKSSSPKSP 1079
QY 310 APTTPKEAPTTKEPTPTTPKEAPTTKEAPTTKEAPTTKEAPTTKEAPTTKEAPTT 366
DB 1080 KTAGPKE--TPGCVAVAPPEISLPKEPTQVATPESLAASSQKSPKTSVPEKTEPPG 1136
QY 367 ---TTPKEAPTTTKEPSPTPK-----EPAPTTKSAPTTTKEAPTTT--KSAPTPK- 416
DB 1137 GVTAMPLEISAPQAKATVAPKQIPTPEDAVITLAGSPLSPKKASKTAAAPKEAPATPSV 1196
QY 417 -----EPSPTTTKEAPTTTKEAPTTTKEAPTTT--PKE-----PAPTTP 455
DB 1197 GVAIVSEISPSPKKTSKTAAPKENSATLPKKSPTKAAKKEIPATSSSEGVAVPSIS 1256
QY 456 KEAP-----TTKKAPATKEAPATTKETATPTPKLTPPTPEKLAPTTPEK--PA 507
DB 1257 SPPTPAKSGVAVLTPGAGNALAE--SPASPKVKVPTAAAEESTTP-----SPQIKP 1309
QY 508 PTPPEELAPTTPEPTPT-----PEEPATTPKAAAPNTPKEAPTTPE--PAPT 558
DB 1310 VAGPKASATPPSKTTPKTAIVPKETSAPSEGVAVAVLEIPSPKAKTAAPEKTPAPS- 1368
QY 559 PKPEAPTTKEAPTTTKEAPTTTKEAPTTT-----KKAPKE 598
DB 1369 -PGATTATPAQIIPSPRKSCKKAGSKE--TPTTSPREGVTAAPLEIPSSKTSKMSAPKE 1426
QY 599 LAPTTKEPT-----TSDKAPATTPKGTAPTTKEAPATTPEK--PAPTTPKGTAPTT 651
DB 1427 TLVTPSSKKLSQTVGPKETSLGATVAVLEIPSHKKAPKTVDPKQVPLPSPK--DAPT 1485
QY 652 LKEAPATTPKKAPKELAPTTTGGPTSTSDKAPATTPKGTAPTTKEAPATTTPKAPPT 711
DB 1486 LAE-SFSSPKK-APKTAAPSER-VTVPEKPA-TPQKASGTTASKVAVPAETQEVAVS 1541

```
OY 712 TPETP-----PPTTSEVPTPTTK-----EPTTHKSPDESPTBELSAEPTPKALENSPK 760
DB 1542 SREPTTPAATPPVKNPSHHKSTKITLKEAPATLPPSPKSPKSPKAPR--SAPK 1599
OY 761 E-PCVPTTKPPAATKPEMTTAKDTERDLRTTPTTTAAPKMTKETATTTETTESKI 819
DB 1600 EFPASPSPK-PVTT--SLAQTAAPSLQKAPSTTIPKLENLAFAV----LPVSSKSPAPA 1652
OY 820 TATTTVOTSTTTTODTPTFKITTLTKTLTLAKVTTTKTTTITTELMNPEEPKAKKDATN 879
DB 1653 RAASASLSPAAAPQATAKEATTPISCKRAATEPIETSTAPSLGAPKETSE---TSV 1708
OY 880 SKATTPKQKPTKAPKPKPTSTKPKPTPRVKRKPRTPTPKMTSTMBELMPTSRIAEAML 939
DB 1709 SKYLMSP-----PKASSSKRASTIP-----ATLPSLKEASVLS---- 1744
OY 940 QTTTRPNQTPNSKILVEVNPKESEDAGAEGEMP 971
DB 1745 -----PTATSSGKSHISPVS-DACSTGTTP 1770

RESULT 14
OYVEL9 PRELIMINARY; PRT; 2112 AA.
AC OYVEL9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CG4090 PROTEIN.
GN CG4090.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Pnydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.F.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe D., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Adair J.F., Agayani A., An H.-J., Andrews-Plamkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris W.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merklow G., Malsbina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reineert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sultskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
```

```
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003716; AAF5402.1; -.
DR FlyBase; FBgn0038492; CG4090.
DR InterPro; IPR002557; Chitin_binding.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF01607; Chitin_bind_2; 10.
DR SMART; SM00494; ChCBD2; 11.
DR PROSITE; PS00022; EGF_1; 1.
SQ SEQUENCE 2112 AA; 219547 MW; B91018E551A5D36 CRC64;

Query Match 11.1%; Score 756; DB 5; Length 2112;
Best Local Similarity 26.4%; Pred. No. 1,7e-42;
Matches 292; Conservative 107; Mismatches 367; Indels 338; Gaps 47;

OY 21 QVSSQELSCKGRCFESFERGECDCD-----AQCKKYDKC-----CPDY 59
DB 953 QSSSSNOTSSSTQKPKPAKCESEETFLADNENCSKFRVNDNGGCTKVSFTCPPN 1012
OY 60 ESFCAEYKD-NKKNRTRKKPKPPVVDGSGLDNGDFVVTTPDSTGHNKVSTSPKI 118
DB 1013 TLMDPEANSCWHPDQIQTKPLCKKVVSGGSSN-----STSSSSSSNNGSSSNG 1066
OY 119 TTKAPINPRLPNSTKETSITVVKETTV---ETKETTTTKQTS---TGKAKTY 171
DB 1067 SSSS-----SSSSNSGSSSTGSSNSGASSGSSGSSGSSNSGSSNSGSSNSTS 1120
OY 172 SAKETOST-----EKTSADLAPTSSKVLAKPYPKAEITTKGP----- 208
DB 1121 SSTSSSSSSNNNGSSSSSSSSSSSTSKPNSECEKYNAGOTGRDCAKRYRCVDN 1180
OY 209 -----ALTPKEPTPTPKPEASTTPEPT 233
DB 1181 DRGFGNVPFSCGPGTWDQMOQACNHMAVKECGGIAIAPPTTSPTTSR-PTTASTSRPS 1239
OY 234 PTTKSAPTPKPEAPTTTSAPTTPKEAPTTTKPEAPTTTKPEAPTTT 282
DB 1240 DQTSSTSRPTGP---PTTARVTARPTTSSEPTTANSSQTSPTVQAPNTDCKRSEFMA 1295
OY 283 -----TKEP---APT-----TTKSAPTPKPEAP 303
DB 1296 DPNNCSFYRCVNRNKGFSIPQCGAGYWDODLOTCHNHNFNNGCTGHESTPKRPPC- 1354
OY 304 TTPKPPAPPTPKPEAPTTKEPPTTPKEAPTTKEAPTTKEAPPTAARKKAPPTPKE 363
DB 1355 -EPATNGTATSTSTSTTP--PPTTDLPLPTSTGLP-PTTTTELPTT-----TTTDL 1403
OY 364 PAPPTPKPEAPTTTKESPPTTPKPEAPTTTKSATTTKKEAPTTTKSAPTTPK---EPSP 420
DB 1404 PPTTTTLPLPTTTSLPPTTTTGLPPTTTGAOPTTTLSSETSTVTTSPBSTPPPS 1463
OY 421 TTPKEAP-----TTPK----- 432
DB 1464 TTTMKRPLPAGTECGSEGYMADPEDCRKYRCINNGASYRKYNFTCPRGIGTGMNEBYQCDY 1523
OY 433 -EPAP---TTPKKAPATP---KEAPTTPKPE-PAPTTTKKAPATVAKPEAPTTPKETA 483
DB 1524 VENIPRCSKLPABEITTTTPEESKDPSTPOGDEPTTAKPI-TKPT-EESTEPKOR 1581
OY 484 PTPPKKTLPTTPELATPTTPEKAPPTPEELAPPTPEEPPTTPEEAPPTPKAAANTP 543
DB 1582 PTTQYPEKPTTTE--PEKQKPL-TTEYQKPLTTEEPPEKQKPL-TTEYQKPTTT 1637
OY 544 KEAPATPKPEAPPT-KEAPATTPKETA-----PTTPK--GTAPTTLKEAPATTPKPA 595
DB 1638 EEPPEKQKPTTTEYQKPTTEPTTTSIPGNVPTTTSVGNVPTTPIPVETTTSPG 1697
OY 596 PKELAPTTKEPTSTSDKAPATTPKGAATTPKEAPTTPKPE-PAPTTPKGAATPLKE 654
DB 1698 YK---PTTTEGPTTTT-LPSTTTDAIOEPTTSKKPEPTTTTESPESSTEGSVTLOPE 1753
```

```

QY 655 PAPTPPK-----PAPKELA-----PTTKGPTSTSDKPAPTTPKETAPTTKEBPAPT 703
Db 1754 POPNVNCSSEGFPPDPECSRYRCVDAKNKGQYAVAFNGCKGTWDTETECNADQV 1813
QY 704 TPKKPAPTTPTPTTSEVSTPTTKEPTTIHKSPDESPPELSAEPPTPALENSPREP 763
Db 1814 SGN-----CSSGQTTTPTTPTTPTTPTTSSGKP-----ETTSKAPEN----- 1852
QY 764 VPTTKPAATKPEMTTAKDKTTERDLRTTPTTTAAPKMTKETATTEKTSKITATT 823
Db 1853 ---TTTWA-----PETTTT-----SSPETTTT---VASETTTTSGT-----TT 1885
QY 824 TQVSTTODTTPKKTTLTKTLTAPKVTTTKTTTTEIMNKPEETANPKDRATNSKAT 883
Db 1886 TAPETTKTKPKP-----ETTTIAGEETSTSKSPTTT----- 1918
QY 884 TPKROKPTKAKKPTSTKPKTMP 907
Db 1919 SPAPSTWTSAP-----CPETGP 1935

RESULT 15
Q9XDH2
ID Q9XDH2 PRELIMINARY; PRT: 763 AA.
AC Q9XDH2;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE PROLINE-RICH MUCIN HOMOLOG.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Espitla C., Lacleite J.L., Mondragon-Palomino M., Martens A.,
RA Zhang Y., Moreno C., Singh M.;
RT "Cloning and characterization of a new member of the PGCS family that
RT is a useful marker of polymorphism in Mycobacterium tuberculosis.";
RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF071081; AAD41594.1;
DR InterPro: IPR002951; Atrophin.
DR InterPro: IPR002965; P-rich_extensin.
DR InterPro: IPR003882; Pistil_extensin.
DR PRINTS: PRO1222; ATROPHIN.
DR PRINTS: PRO1217; PRICHEXTENSIN.
DR PRINTS: PRO1218; PSTLEXTENSIN.
SO SEQUENCE 763 AA; 75034 MW; 39168EC45A5916F8 CRC64;

```

Query Match 10.5%; Score 715; DB 2; Length 763;
 Best Local Similarity 28.2%; Pred. No. 3.4e-40;

Matches 248; Conservative 39; Mismatches 361; Indels 232; Gaps 39;

```

QY 197 PTPPAETTTKGPALTTTKEPTTTPPKPASTTKEPTTIKSAPTTKEBPAPTTRKSAP 256
Db 3 PVP-----APRALALPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 46
QY 257 TTPKEPAPTTPPKPAPTTKEPAPTTPPKPAPTTKEPAPTTPPKPAPTTKEPAPTTPKE 316
Db 47 PCP---PAP-----PAPKPKSKAPFPVPPAPAPARELAPLP---PAP-----PEAPRE 90
QY 317 PAPTPKEPTPTT-----PKAPATTEKPAFTTPK-EPAPAPK-KPAPTTPKEPAPTTPK 370
Db 91 SRPALPCPPPPVIVIPPPAPAPVPAVNSPPFPFPAPKFAVAPVPP---PVPNSPP 148
QY 371 EP---APTTPKESPTTPKEPAPTTPKSAPTTKEPAPTTPKSAPTTPKESPTTPKEPAP 428
Db 149 FPPPPAPALNPPAP-----PAPPLANSPLPAPAPPTPAGT---PPAPAMPVPPAPAPSKPA 201
QY 429 TTPKEPAPTTPKPAPTTKEPAPTTPKEPAP-----TTTKKPAPTAPKPAPTTKEPAP 484

```

```

Db 202 SPREPAP-----PMPATIMEPPPLPVPVPPDISKETPPAPAPAPPIPPAPVIPPVPLP 256
QY 485 TTPKLTPTTPEKLPAP-----TTPKEPAPTTPPEELAPTTPEEPPTTPPEBPAPTTP-- 534
Db 257 PVPKNIIPAPP---APVAVAAVLVAPCPPLPPLPNHPPAPAPAPVPPVAPVAPLPNSHP 313
QY 535 ---PKAAPNTPKEBPAP----- 548
Db 314 PAPSAVPVGVPLAPLPISGRVSVWKGSTTTSTFCCRCVSGGVLAGALNPSRPSPL 373
QY 549 --TTPKEPAPTTPKEPAPTTPKEA---PTPKGTA---PTLKEPAPTTP-KKPAKPE 598
Db 374 TTTTPALDAPLPPLPLPLPINTAVPPIPLPLPEVTAALAPLPPLAPLPISGVAPAP-- 431
QY 599 LAPTTKEPTSTSDKPAPTTPKGTAPTTPKEPA-----PTTPKEPAPTTPPGTAPTL 652
Db 432 --PAPGKPMWTPPLAPAPPEPK-TVPVLPGGSCPPSEKPNPAPPEPEPKSSDALP 488
QY 653 KEPAPTTP---KKPAKELAPTTKGP-TSTSDKPAPTTPKET---APTTPKEPAPTTP 705
Db 489 APAPSMPSAVRVPSPPIPPAPAPAPASMPALPAPAPSPATRLCPPLPSPAPAPNSP 548
QY 706 KKPAPTTPETPTTSEVSTPTTKEPTTIHKSP-----DESTPELSAEPTPKA--LENSP 759
Db 549 --PAPAPAPPPPKILS--ANPPCPVPVPPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 604
QY 760 KEGVPTTKPPAATKPEMTTAKDKTTERDLRTTPTTAAAPKMTKETATTEKTESKI 819
Db 605 --PAPAPAPAPPSALPFVNPAP-----PPTPAAR-----SRPAL 637
QY 820 TATTTQVSTTODTTPPKTTTLTKTLTAPKVTTTKTTTTEIMNKPEETAKPRDRATN 879
Db 638 PAAPAPAPAPVATTP----- 880
QY 880 SKATTPKPKP---TKAPKKTSTKPKTMPRVKPKTTPPRKMTSM-----PELNP 930
Db 666 SMALPAPDPDPILATPAPAPAPPLPMSPPAPPLPAPAPAPAPAPPLITINOPSPPLAP 725
QY 931 TSLIAAMLOTTTRPQOTNSKLVEYNPKSEDAAGAEGET 970
Db 726 VEGAPLAPLPINGRPVFAKRNLSI-----GSSSGDT 756

```

Search completed: April 26, 2002, 16:29:32
 Job time: 587 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:20:05 ; Search time 114.61 seconds
(without alignments)
816.844 Million cell updates/sec

Title: AAI
Perfect score: 6568
Sequence: 1 MAWKTLPIYLLLSVEFIQ.....ARATTRSGQTLKSWYNCP 1229

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1062.5	16.2	3020	2 A43932	muscin 2 precursor,
2	950	14.5	1664	2 T18262	S-layer protein -
3	853	13.0	1489	2 T31108	cyst germination s
4	802	12.2	1274	2 T16251	hypothetical prote
5	789.5	12.0	1367	1 S48478	glucan 1,4-alpha-g
6	784	11.9	1188	2 S49915	extensin-like prot
7	778.5	11.9	2187	2 T30826	nascent polyphid
8	672	10.2	1344	1 A35175	muscin 1 precursor,
9	671	10.2	1151	2 T18535	high molecular mas
10	659.5	10.0	1229	2 T25697	hypothetical prote
11	635	9.7	7962	2 T18346	elastin class 1 p
12	633	9.6	489	2 T11622	elastin class 1 p
13	632	9.6	761	2 C84672	hypothetical prote
14	629	9.6	3507	2 T34513	hypothetical prote
15	627	9.5	990	2 T51618	nucleolar phosphop
16	625	9.5	6642	2 T29757	protein UNC-89 - C
17	622.5	9.5	839	2 T75518	hypothetical prote
18	622	9.5	3570	2 T45025	muscin MUC5B, trach
19	620.5	9.4	971	2 T19431	hypothetical prote
20	607.5	9.2	801	2 T29018	hypothetical prote
21	607.5	9.2	924	2 S27923	gene LF3 protein -
22	605	9.2	379	2 S50125	larval glue protei
23	562	8.6	1630	2 A53577	ascites sialoglyco
24	559.5	8.5	350	2 S22456	hydroxyproline-ric
25	556.5	8.5	856	2 T16543	hypothetical prote
26	556.5	8.5	875	2 S23760	polypheolic adhes
27	553	8.4	2232	2 T34434	hypothetical prote
28	551	8.4	620	2 S06733	hydroxyproline-ric
29	545	8.3	873	2 A47283	calphostin - fruit

30	542	8.3	369	2 S20500	hydroxyproline-ric
31	538.5	8.2	416	2 J00465	extensin precursor
32	533	8.1	1087	1 OFMSH	neurofilament trip
33	532	8.1	756	2 T37642	hypothetical prote
34	530.5	8.1	1162	2 JH0557	exo-alpha-sialidas
35	530	8.1	865	2 A47282	calcium-binding pr
36	522	7.9	328	2 J00985	hydroxyproline-ric
37	522	7.9	1459	2 T32271	hypothetical prote
38	518	7.9	813	2 S70795	vsaa protein precu
39	518	7.9	866	2 T45462	membrane glycoprot
40	518	7.9	1072	1 A37221	neurofilament trip
41	512.5	7.8	867	2 T45463	membrane glycoprot
42	509.5	7.8	662	2 A45155	muscin FIM-C.1 - Af
43	508	7.7	1832	2 T31113	muscin-like glycopr
44	506.5	7.7	700	2 A54641	interspersed repea
45	504.5	7.7	606	2 A43427	neurofilament trip

ALIGNMENTS

RESULT 1
A43932
Muscin 2 precursor, intestinal - human (fragments)
N:Alternate names: muscin SMDC-41
C:Species: Homo sapiens (man)
C:Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 05-Nov-1999
C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal muscin (MUC2) cDNA. Identification of t
A:Reference number: A49963; MUID:94132002
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
A:Cross-References: GB:L21998
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal muscin has cysteine-rich subdomains located both up
A:Reference number: A45106; MUID:93016075
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Residues: 626-1895 <GU2>
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-References: GB:M94131; NID:G186395; PIDN:AA59163.1; PID:G186396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-References: GB:M94132; NID:G186397; PIDN:AA59164.1; PID:G186398
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal muscin gene structure. Repeated arrays and polym
A:Reference number: A43932; MUID:91358717
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-References: GB:M74027; NID:G188863; PIDN:AA59875.1; PID:G188864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal muscin cDNAs. Sequence analysis and evl
A:Reference number: A33532; MUID:89197956
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GUA>
A:Cross-References: GB:M22405; NID:G188873; PIDN:AA36334.1; PID:G188874
A:Experimental source: Intestine
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; MUID:91086481
 A:Accession: A61257
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <TAN>
 A:Experimental source: bronchus
 Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus
 A:Reference number: P00328; MUID:92198477
 A:Accession: P00328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M86523
 A:Experimental source: small intestine
 A:Accession: P00329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 C:Genetics:
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:120203; OMIM:158370
 A:Map position: 11p15.3-11p15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von Willebrand factor type C repeat homology
 C:Keywords: glycoprotein; intestine; tandem repeat
 F:2766-2834/Domains: von Willebrand factor type C repeat homology <WMC>

Query Match 16.2%; Score 1062.5; DB 2; Length 3020;
 Best Local Similarity 29.2%; Pred. No. 8.3e-40;
 Matches 342; Conservative 74; Mismatches 442; Indels 315; Gaps 39;

56 FKTTPTDSTGTH---NKVSTSPKITAKINRPSLPPSDTSKESLSLVNKEVETK 112
 DB 1240 WEICGPGVIEKHNICISITRPSLTFTTTLPTTP-----TSFTTTTTPIS 1291
 QY 113 ET---TTTNK-----QSTDSKEKTS--AKETOSIEKTSKD----- 145
 DB 1292 SYLSTTPKLCGLMSDMINEDHPSGSDGDRPFPGVCAPDIEDRSKVDHLSIEQ 1351
 QY 146 -----LAFTSKVLAKPPPKAE 161
 DB 1352 GOKVOCDSVGLCKNEDQNGPGLCYDKIRVNCWPMDCITTPSPPTTSPSPPT 1411
 QY 162 TTTKGALTTPKPEPTTTPKEPASTTPKEPTTIKSAP--TPKEPAPTTPKSAPTKE 220
 DB 1412 TTTTLPTTTPSPPTTTTTPPTTTPSPPTTTTTPPLPLTPSPSTTTPPTTTPS 1471
 QY 221 PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAP--TPKEPAPTTPKEPAPTTP 279
 DB 1472 P-PTTTPSPPTTTPSPPT--TTTTPTTTPSPDMWTPIIPASSTTLPLPTTTPSPPTT 1529
 QY 280 TPKEPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKAPAPTTPKEPAPTTPKEP 339
 DB 1530 TTPPTTPPTTTPSPPTTTPP-PTSTTLPLPTTTPSPPTTTPPTTTPSPPTTTPSP 1588
 QY 340 SPTTPKEPAPTTPKSAP--TTTKEPAPTTPKSAP--TPKEPAPTTPKEPAPTTPKEPAPT 397
 DB 1589 PTITTTTPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPITPPTSTTLPLPTTTPSPPTTTP 1648
 QY 398 PKKPAPTTPKEPAPTTPKEPAPTTPKAPAPTTPKAPAPTTPKEPAPTTPKEPAPTTP 457
 DB 1649 TTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSSPTITTP--SPPTT--TMTTPS-- 1700
 QY 458 APPTPKAPAPTTPPELAPPTPEEPTTPTEEPAPTTPKAAAPNTPKAPAPTTPKEPAPTTP 517
 DB 1701 PTTTPSSPTTTPPTTTPSPPTTTP--SPPTTMTTSPPTTTPSPPTTTPMTLPLPTTTPSSPLTTT 1759
 QY 518 PKEPAPTTP-----KETAPTTP-----KG----- 536
 DB 1760 PLPPTTTPPTTTPSPPTTTPPTTTPPCVPLCNMTGWLDSGKPNFHKRGDTTELIGDVCGPGMMA 1819

QY 537 ----- 536
 DB 1820 NISORATMYDVPITGLQGVVCDVSGVGLCKNEDQKRGGVITMAPCLNTEINVOCECV 1879
 QY 537 TAPTTKEPAPTTPKAPKELAPT--TKEPTSTSDK-----APT--TPKGAAPTTPKE 569
 DB 1880 TQPTTM---TTTTVENPTTPITTTTTLVPTTPPTSPSTQSGNGLQAPPTPPTSTSTTTP 1936
 QY 590 PAPTTPKEPAPTTPKAPAPTTPKEPAPTTPKAPKELAPTTPKAPAPTTPKAPAPTTP 648
 DB 1937 PTPPTGTPP--TTPITTTTTPPTTPPTGTPPTVLTITTTTTPPTTPPTSTSTTTP 1995
 QY 649 KETAPTTPKEPAPTTPKAP-----APTTPPTTPPTSPSTSTTTPK-----EPTTHK 696
 DB 1996 ITTTTTPATPTPTGTPPTPIMISTTTTTPPTTPPTTPGTPPTSTGPTHTSTAPIELTSPN 2055
 QY 697 SPDSPTPELSAEPPTPKALENSPKPEPVPTTKTPAATPKEMATTAKODTERDITTPET 756
 DB 2056 PPESTTPQTSRSTSSPPTPESTTTLSTLPPAIEMTSTAPSTPAPPTTSGHTLSPSPST 2115
 QY 757 TAPKMTKETATTTTEKTESKITATTOVSTTQDTTPKIT--LKTTLAPAVTTP 813
 DB 2116 TTSPEGTPTRGTT--GSSAPPTSTVQTTTTSAMTPTTPPLSTPLIRTTGLRPSSV 2173
 QY 814 KTTITTEINKKPEE-----TAKPKDRATNSKATTP 844
 DB 2174 LICCVLNDITYAAGEEYVNGTGYDTCFVNCSLSEFLFNMSCPSTPSTTPPSK--STP 2232
 QY 845 KPGKPTAPKPKPTSTKPKTTPRVKRPKTTTPPRKMTSTPELNPSTRIAEAMLQTTTP 904
 DB 2233 TPSPKSTPSPKPPGKTPPCPDPPR-----QNETFMWLCDCFM--ATCKY 2278
 QY 905 NQPNKSLVEVP-----KSDAG-----AEGTFPHLL----- 934
 DB 2279 NNTVEIVKCEPPEPMPETCSNGIQPVAVEPDDCCMHMECDYCTGWD--PHVTPDGLX 2337
 QY 935 -----RPHVPEVTPDMDYLPVRVNOGILIN 961
 DB 2338 YSYQNGCTVYLVEISPSVD-----NFGYIID 2364

RESULT 2
 T18262
 S-layer protein - Clostridium thermocellum
 C:Species: Clostridium thermocellum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18262
 R:Fujino, T.; Beguhn, P.; Aubert, J. P.
 J. Bacteriol. 175, 1891-1899, 1993
 A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose
 e.
 A:Reference number: Z18847; MUID:93209931
 A:Accession: T18262
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1664 <FUU>
 A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match 14.5%; Score 950; DB 2; Length 1664;
 Best Local Similarity 31.0%; Pred. No. 4.4e-35;
 Matches 312; Conservative 104; Mismatches 361; Indels 230; Gaps 55;

QY 152 VLAKPTP-KAETTTKCPALTTKEPPTTPKEPASTTPKEPPTTTPKSAPTTPKEPAPTTP 210
 DB 758 VLIQAPITKAASDEPIPTDPSDEPTP-----DEPTPS--DEPTSDDEPTSD 804
 QY 211 TKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 266
 DB 805 EPITPSTPEPTPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPTDTP 864
 QY 267 KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 325

Db 865 SDEPTPSDEPTPS--DEPTPS--DEPTPS--SDEPTSETEEPPIPTDPSDEPTPSDEPT 919
 QY 326 TTPKEPA--PTTKEPSPT--TPKEPAITTTKS--APTTPKEAPT--TTKSAPTTPKEPS 378
 Db 920 SDEPTPSDEPTPSDEPTSETEEPPIPTDPSDEPTPSDEPTPSDEPTPSDEPT 979
 QY 379 ---PTTKEAPT--TPKPAAPTTPKKAAPTTPKEAPTTPKKAAPTTPKKAAPTTPKKAAPT 433
 Db 980 PSDEPTPSDEPTPSDEPTSETEEPPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1039
 QY 434 APT---TPKEAPTTPKKAAPTTPKEAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 486
 Db 1040 TSDSEPTSE---TPKEAPTTPDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1094
 QY 487 EBPAPTTPKKAAPTTPKEAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 543
 Db 1095 EBPAPTTPDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1154
 QY 544 EBPAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 602
 Db 1155 EBPAPTTPDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1208
 QY 603 PGTAAPTTPKEAPT--TPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 658
 Db 1209 ---DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1253
 QY 659 PAPTTPKKAAPT--TPETPTPTSEVSTPTTPKEAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 717
 Db 1254 PAPS--DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1305
 QY 718 PKEPVPTTKPAAT--KEMTTAKDKTTENDLRTPTETTPAAPTTPKKAAPTTPKKAAPTTPKKAAPT 775
 Db 1306 SDEPTPSDEPTPSDEPTSETEEPPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1361
 QY 776 SKITATTTQVSTTQDTPTFKITTKTTTAPKVTTKTTTKTTTKTTTKTTTKTTTKTTTKTTTKTT 835
 Db 1362 DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1396
 QY 836 AATNSKAT-----TPKQKPTKAP--KPTSTKPKKPTAPRVKPKPTPTPKPM----- 880
 Db 1397 SGGSGSGSGGGGGGGGTVPSTPTPTSKPTSPAP--TEIEPTPSDVPKALIGEENRA 1453
 QY 881 -----TSTMBELNPTSTIAEAMLOTTTRPNOTPNSKIVEVNPKSEDAAGAGETPHM- 932
 Db 1454 YLRGYPDGSFRERNITRAEAAVIF-----AKLL---GADESYGAQASAPSYD 1498
 QY 933 LARPH-----VEMPE-----VTPOMDLPRVPPNGI----- 958
 Db 1499 LADTHAANAIAIFANSQGLFKGYPDGTFRPDONTTRAERATVLAHLTVKGOELMSKLA 1558
 QY 959 ---IINPMLSDETNICNG---KPVDELTLT---RNGTLVAFRGHYFMMLSPPSPS 1005
 Db 1559 TIDISNPKDD---CVGMWQEFIKTLGLGISGYPGT-----FKPON 1600
 QY 1006 PARRTIEWGIPSPIDVTFRNCNCEKTEFFKD--SOYRFTDIDK 1050
 Db 1601 YIKRSESV---ALINRALEGRPLNGAPKLFDPVNESYAF--GDIMD 1642
 RESULT 3
 T31108
 cyst germination specific acidic repeat protein precursor - Phytophthora infestans
 C:Species: Phytophthora infestans (potato late blight agent)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T31108
 R:Goenhardt, B.
 Submitted to the EMBL Data Library, April 1998
 A:Reference number: 220986
 A:Accession: T31108
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1489 <GOE>
 A:Cross-references: EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAC72308.1

C:Genetics:
 A:Gene: car90
 Query Match 13.0%; Score 853; DB 2; Length 1489;
 Best Local Similarity 32.3%; Pred. No. 7,9e-31;
 Matches 355; Conservative 46; Mismatches 475; Indels 222; Gaps 49;
 QY 33 TTKKPKPPVVDAGSGDNGDFKVTPTDTSST-----OHKKVSTSPKITAKINRA 86
 Db 338 TTPAPTEKDYDEETTYVVEESTYAPTKSETNAPTRMHAHLEKCDIEVMYAPTEET 397
 QY 87 PSLP-----PNSDT---SKESLVNKKETVETKEETT--TNKQTSND 124
 Db 398 TYAPTEETTYAPTEETTYAPTEETPYEPTTEETTYPTTEETTYAPTEETTYAPTEETTYAP 457
 QY 125 GKEKTSAAETOSIEKTSKDLAPTSKVLAKPRPKAETTTKGALTPKEAPTTPKKAAPT 184
 Db 458 TEETTYAPTEETPYEPTTEETTYAPKETTYPAT--EETTYASTEETTYAPTEETTYAPA 514
 QY 185 STTPKEPTPTTIKSAPTTPKEAPT--TKSAPTTPKEAPT--TPKKAAPTTPKKAAPTTPKKAAPT 241
 Db 515 EETPYEPTTEET--TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 573
 QY 242 TKEAPATTTKSAPT-----TPKEAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 291
 Db 574 TYAPTEET--YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 632
 QY 292 P-----AP---TPKEAPTTPKEP-----APAPKKAAPTTPKKAAPTTPKKAAPT 328
 Db 633 PTEETTYASTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEET 692
 QY 329 KEAPAT--TPKEAPTTPKEP-----APT--TKKSAPT-----TPKKAAPTTPKKAAPTTPKKAAPT 363
 Db 693 TTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 752
 QY 364 PT--TTKSAPT--TPKPSPTTKKEP-----APTTPKEAPTTPKKAAPTTPKKAAPTTPKKAAPT 412
 Db 753 PTEATYAPTEETTYAPAEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 812
 QY 413 TPKEAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 472
 Db 813 TPKEAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 870
 QY 473 LAPT--TPKEAPTTPKEP-----APT-----TPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 513
 Db 871 YAPTEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 930
 QY 514 -----APTTPKEAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 563
 Db 931 EETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 990
 QY 564 KEPTSTTSKPAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 609
 Db 991 YAPTEETMYAPTEETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEET 1050
 QY 610 -----TPKEAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 648
 Db 1051 EETTYAPTEETTYAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1110
 QY 649 KETAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 706
 Db 1111 YAPAEETPYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1169
 QY 707 AETTPKA--LENSPKPG-----VPTTKTPAALKPKEMTTAKDKTTRDLRTPT---ETT 756
 Db 1170 TEETPYAPTEETTYEPTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1229
 QY 757 TAAPKMTKETATTTTEKTSKITATTOVSTTQDT-----PEKITT---LMT 803
 Db 1230 TYAP-----TEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET 1284
 QY 804 TPLAPKVTTKKTTTTEIMNKPTEETAKPKDRAT--NSKATPKP--OKPTKAPKPKPT--- 857

```

Db 1285 TTYAPATATYAPTEEDPPVAPTEETTEGTYAPTEETTYAPTEETTYAPMEETPYE 1344
Oy 858 -----STKRP-KTMPVVRKKTTPTRKMTSTMPELNPTSRILAEAMLOTTTRP--N 905
Db 1345 PAESTSTVSTKRCNTEETFTDEPTDEPTDE--PSDEPTDEPTDEPTDEPTDEPTDEPCDN 1402
Oy 906 QTPNSKLVENVPKSEDA 923
Db 1403 QGINGIGVENKVKYNNAG 1420

RESULT 4
116251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T16251
R:Leinbach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: Z18485
A:Accession: T16251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1274 <LEI>
A:Cross-references: EMBL:U06675; NID:g1166613; PID:g1166621; PIDN:AAB52641.1; GSPDB:GNOC
A:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP:F35A5.1
A:Map position: X
A:Introns: 1272/2

```

```

Query Match 12.28; Score 802; DB 2; Length 1274;
Best Local Similarity 28.0%; Pred. No. 1.2e-28;
Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

```

```

Oy 30 KNRTKK-KP-----TP-----KPPVDEAGSGIDND-----FRVTPTDSTTQHNKYSTS 74
Db 274 KNRTKKRPWEDETPVEEKEPPVPEKKAPVLKKDPAPAAKADSPSKAPKVEPS 333
Oy 75 PKITTAKP-I-NRPSLS-PPNSDTSKETSILYVKEETVETKTTTNGKSTDGKEKITSAK 133
Db 334 SPVVPPTPVKNPKYKKYKPEVDEDEPAEYKKSAPPEKTPVLKRRKEPSSTTPSDDPS 393
Oy 134 ETOSIEKTSKDLAPTSKVLAKPTPKAE-----TTTGPA-----LTTT 172
Db 394 PKKAAVAVKPRDSSPKKAPPLQADPKQEVPPVKNPVKVKYKPPWEVDEDEPAEYKOP 453
Oy 173 KEPTPTTP-----KEPASTTPKEP-----TP--TTIKSAPTTKEPAPTTTGSAPTTKPEPAP 223
Db 454 EAAKAKTTPVLKREKPAKADAKPATSKTPTPEKDPVKKRDSPPKVAAPSAQAPPA- 512
Oy 224 TTTKEPA-----PTTPKEPAPTTTKEPAPTT-----TTKSAPTTT 257
Db 513 TPVKNPVKMKRPMWEDETPADVDVSKPTQAKTTPSLAKDPAKESLKKADTKAPAKP 572
Oy 258 KEP-----ATTPKKAPPTTPKE-----PAPTTPKEPTTPPTTPKEPAPTTTKEPAPTTTKEP-EP 307
Db 573 RDSPPKVPAPTAPKEKTPVLAKKEPAGPADSKTEPEKSPKRPSPSKAVAPKVPETTEV 632
Oy 308 APAPKPPAP-----TTTKEPAPTTTKEPAPTTTKEP-----SPTTPKEPAPTTTTSAPT 357
Db 633 APAAVKKEPISKPDAPKKAEPNSFYVP-PTPVKNPVKMKRPMWEDEDDAPAKPVSLPE 691
Oy 358 TTKEPAPTTTTSAPTTT-KEPSEPTTKEPAPTTTTPK--EPAPTTTTPKVPAP--TTTPKEPAPT 412
Db 692 PEKK-TPVLAKKAPPTKPDSDAADPVSGSPSKDPLAKKAPVAPRDSPPKAVPIKAPK 750
Oy 413 TTKEPAPTTTTPKAPPA--PEEPAPTTTPKEPAPTTTPKKTTPTT-----EKLAPTTPEKPA 466
Db 751 T--EVPAPVVKKEPEVAKSRDPSPKAK-AEPNSP--VVPPTPVKNPVKMKRPMWEDDDA 805

```

```

Oy 467 PTTPEELAPTTPEEP-----TPTTPEEPAPTTPKAAAPNTPKCBAPTTPEKAPATT--PK 520
Db 806 PAEIVANPEPEKKTIVLAKKTPVKKRDSPPKAVPAKPSKTDAPVSVKKEPVSKPE 865
Oy 521 PAPTTPKEPAPTTTPKGTAPTTTLEKAPATTTPKKAPKAPKAPAT--TTKEPTSTTSKPAPTTP 579
Db 866 PSPKKAEPPNSPVVP-----PTPVKNPVKMK--KPPWEDEDETEEVKKRSE--PEKKTPVLA 918
Oy 580 KGTAPTTPEKAPATTTPKEPAPTTTPKGTAPTTTLEKAPAT-----TPKKAP-----KEL 627
Db 919 K-KEPEKPD-APKVAAPRDPSPKAVPE--KEPAKVAAPRDLSEKKAIPAPANTOEA 974
Oy 628 APPTTKGPTS-----TTSKP-----APTPKET-----APTTPEEPAPTTTPKAPAT 670
Db 975 PTPPVKNPVKMKRPMWEDEDEPAEVPASAPPEKKTIVLAKKAPKAPRDP--SPKKAAPV 1031
Oy 671 TPTTPPTTSEVSTPTTPKPTTIHKS-----DESTPELSA-EP--TPKALENSKPEPG 722
Db 1032 AAK-PDCKIPEV-PTTPVKNPVKMKRPMWEDEDESPSPASAPPEKKTIVLAKKAPTKPA 1089
Oy 723 V-----PTTKPAAAT-----KPEMTTAKOKTTERDLRTTPETT--TAAPK 761
Db 1090 TKPDSEAAADPVSGPITSKDKLSKAPVEKPKPTTDPKDKLKPSPAKKPEKAPAPAPK 1149
Oy 762 MKRETAATTTTKEKTESKTTATTTQVSTTQDTTPKITTILKTTTLAPKVTTKKTTTTE 821
Db 1150 KMKPVWDDEDEPADTTPVAPSKKPTEDPADPLG-----GPKTDPK----- 1193
Oy 822 IMNKPEETAKPDRAINSKATTPPKOKPTAPKPKPTSTKKPKTMPVVRKPK----- 872
Db 1194 -LNKKAAPAEKPTER-----PKPKEVSKPEPKPTPEPKP-AAAPKMKRPMWEDEDEPE 1243
Oy 873 ---TTPPKMTSTMP 885
Db 1244 ADETPMAPKPKRDEDP 1259

```

```

RESULT 5
S48478
glican 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein Y1R01
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C:Accession: S48478; B26877; S27281; J06123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48478
A:Molecule type: DNA
A:Residues: 1-1367 <ROW>
A:Cross-references: GB:Z47047; EMBL:Z38061; NID:9603997; PID:g763364; GSPDB:GN00009;
R:Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A:Reference number: A91831; MUID:87194600
A:Accession: A26877
A:Molecule type: DNA
A:Residues: 1-242 <YAM>
A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A:Accession: B26877
A:Molecule type: DNA
A:Residues: 762-1331 <A2>
A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R:Parodo, J.M.; Ianez, E.; Zalacain, M.; Clarios, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacc
A:Reference number: S27281; MUID:89031230
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 1-31 <PAR>
A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CMA32069.1; PID:g4552
R:Iambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.

```

Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Mucin, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A:Reference number: JC6123; MUID:9632327
A:Accession: JC6123
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <LAW>
A:Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C:Genetics:
A:Gene: SGD:MUC1; SPA2; MAL5; DEX2; SGD:S0001458
A:Cross-references: MIPS:YR019C; SGD:S0001458
A:Map position: 9R
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F:5-21/Domain: transmembrane #status predicted <TM>
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 12.0%; Score 789.5; DB 1; Length 1367;
Best Local Similarity 28.4%; Pred. No. 4.7e-28;
Matches 301; Conservative 104; Mismatches 483; Indels 171; Gaps 47;

```
QY 59 TTPDSTTQHKNVSTSPKTTAKPINP-----RPSLPNDSKTSLTVNKEITVE 110
Db 236 STSESTTSSTSESSSTTAPATPTTSCTKERPTPTTSCTKERPTPPHHDTTPC 295
QY 111 TKETTTNNKQTSDDGKEKTSKAKTOSIEKTSKADLAPTSKVLAKPTPKAETT--TKGPA 168
Db 296 TKKTTTSK-TCT--KKTTTPVPTPS-SSTESSAPV-----PTPSSSTESSAPV 344
QY 169 LTPKE-----TTPTPKEAPTTPKEPTTTIKSAP-----TTPKEAPTTPKSAPTTP 218
Db 345 TSTSTESSAPVPTPSSSTESSAPVSTSTESSAPVSTSTESSAPVPTPSSSTTE 404
QY 219 KEAPTTPKEAPTTPKEAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKEAP 278
Db 405 SSAPVTSS-----TTSSAPVTS-----STESSAPVT-----SSTESSAPVTSSTTES 453
QY 279 TTPKEPTP---TTPKEAPTTPKEAPTTPKEAPTTPKAPTTPKEAP---TTPKEP 331
Db 454 SSAPVPTPSSSTESSAPVT---SSTESSAP-VPTPSSSTESSAPVSTSTESS 509
QY 332 APTTPKEPTTPKEAPTTPKSAPTTPKEAP-----TTKSAPTTPKEPTTPKEAP 387
Db 510 APVPTPS-SSTESSAPAPTTPSSSTESSAPVSTSTESSAPVPTPSSSTESSSTP 568
QY 388 ----TTPKEAPT-TPKKAPTTPKEAPTTPKEAPTTPKAPTTPKAPTTPKAPTTPKEA 442
Db 569 VTSSTESSAPVPTPSSSTESSAPVT-----PSSSTESSAPAPTTPSSSTESSA 624
QY 443 PTPPKLTPPTPEKLAPTPKEAPTTPPELAPT-TPEBPTPTTPEEAP-----TTPKA 496
Db 625 PVT-----SSTESSAP-VPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESS 679
QY 497 AAPNTPEAPTTPKEAPTTPKEAPTTPKETAPT-TPKGTAPTLKEAPTTPKAP 555
Db 680 SAPVT-----SSTESSAPVT-----SSTESSAPVPTPSSSTESSAPVPTPSSSTTE 731
QY 556 KELAPTTKEPTSTSDKAPTTPKGTAPTTPKEAP---TTPKEAPT-TPKGTAPTTL 611
Db 732 SSAPVPTPS-SSTESSAPVTSSTESSAPVPTPSSSTESSAPVPTPSSSTESS 790
QY 612 KEAPTTPKAPKAPKELAPTTTKGPTSTSDKAPTTPKETAPTTPKEAP-----TTPKK 666
Db 791 SAPVPTPSSSTESSAPVPTPSSSNITSSAPSTTPPSSSTESSSVPTPSSSTESS 850
QY 667 PAPTTPPTTSEVSPT-----TTKEPTTIHKSDESTBELSAEPTPKALENSPKCP 721
Db 851 SAPVSSSTESSAPVPTPSSSNITSSAPSSI---PSSSTTESFSTGT-TVTPSSSKYP 906
QY 722 GYPTTKTTPAATKEMTTAKOTTERDLRTPPTTTAAPKMTKETATTTTEKTESKITAT 781
Db 907 GSQETSVSSTETIVPTKTTSTVPTSTTTTITTVCSGTNSAGETSGCSPKTVTTT 966
```

```
QY 782 --TTQVSTTTTQDTPPKIITLTKTTLAPKVTT---TKKTIITT-EIMNKPEETAKPKDR 835
Db 967 VPTTTTTSVTSSTTTTITTTVCSTGNSAGETSGCSPKTIITTVPCSTSPSETA----- 1021
QY 836 ATNSKATTPKPKQP-----TKAPKPTSTKKPKTMPRVKPKKTTPTPR 878
Db 1022 ---SESTTSFTPTVTVVSTTVVTEYSTKPGGEITTFVTKNPTTYLTIAPTP- 1077
QY 879 KMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVEYNPKSEADAGGAETPHMLLRPHV 938
Db 1078 -SVTITVTFPTT-ITTVCSGT-----NSAGETSGCSPKTIITTVPCST 1122
QY 939 FMEVPTPMDYLDPRVNOGIIINPLMSDEFINICNRPVGLTLRLNGLTVAFRGHYFWML 998
Db 1123 GTGEYTEATLTAVTTTAVTTTSTESSTGTNSA-GKTTTGYTTKSVPTT-----YVTTL 1175
QY 999 SPFSPSPARRITEVVGIPSPIDTVFTRC-----NCEGKT 1033
Db 1176 APSAPVTPATN-----AVPTTIIT--TECSAATNAGET 1207
```

RESULT 6
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

Query Match 11.9%; Score 784; DB 2; Length 1188;
Best Local Similarity 29.9%; Pred. No. 7.2e-28;
Matches 230; Conservative 59; Mismatches 122; Indels 122; Gaps 33;

```
QY 37 PTPKPPVDEAGSLDNGDFKVTTPDTSTQHNKVSSTPKIITAKPINRPSL----PPN 92
Db 460 PTPHSPAD-----DIVPTTPVPGKSPATSPSQVQPPAASTPPPSLVKLUSPPQ 510
QY 93 SPTSKESTLVNKEETVETKETTNNKQTSDDGKEKTSKAKTOSIEKTSKADLAPTSKV 152
Db 511 APVG-----SPPPPVKTTPPAPIG---SPSPPPPVSVV 541
QY 153 -----LAKTPKAETTTKGALTTPKETPTTPKEPASTTPKEPTTTIKSAPTTKEPA 207
Db 542 SPPPPVKSPPPPAPVGSPPPPPEKSPPPAPVASPPPPVKSP---PPPTLVASPPPPVKSP 599
QY 208 PTTTKSAPTTP--KEAPTTPKEAPTTPKEAPTTPKAPTTPKAPTTPKAPTTPKAPT 264
Db 600 PAPVASSPPPPVKSPPPPTPVASPPPPAPVASPPPPPMKSPPPPTPVSSPPPPPKSPPPP 659
QY 265 PKKAPTTPKEAPTTP-----PKE---PTTTPKEAPTTPKEAPTTPKAPTTPKAPTAPK 315
Db 660 PAKSTPPPEYTPPTTSVKSSPPPEKSLPPTLIPSPPPQEKPTPSTPSKP--PSSPEKP 718
QY 316 APTTPKEAPTTPKEAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPK 375
Db 719 SP--PKEPVSSPPQTP-----KSPPPAPVSSPPPTPVSSPPALAPVSSPPKSSP---- 768
QY 376 EPSPTTTKAPTTPKEAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTAPK 435
Db 769 PPAVLSSPPPAQVKSPPPPVQVSSP-PPAPKSSPPPLAPVSSPPQVEKTSPPAPLSSPP 827
QY 436 TTPKETAP-----TTP-----KLTPTTPEKAPTTPKEKAP-----TTPBELAPTPEPT 482
```


A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA5806.1; PID:g182129; GB:
 A:Experimental source: splice form B
 A:Note: GenBank entries HUMEPIS1B1 and HUMEPIS1B2 present only the amino- and carboxyl-er
 R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Burchel
 J. Biol. Chem. 265, 15286-15293, 1990
 A:Title: Molecular cloning and expression of human tumor-associated polymorphic epithelial
 A:Reference number: A35886; MUID:90368715
 A:Accession: A35886
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-19,29-992,1033-1344 <GEN>
 A:Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870
 A:Note: GenBank entry HUMUCAB includes one copy of the tandemly repeated sequence
 R:Lin, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
 J. Biol. Chem. 265, 15294-15299, 1990
 A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
 A:Reference number: A35887; MUID:90368716
 A:Accession: A35887
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
 A:Cross-references: GB:J05582; NID:g189598; PIDN:AAA60019.1; PID:g189599
 R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky, J.
 Eur. J. Biochem. 189, 463-473, 1990
 A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may genera
 A:Reference number: S10571; MUID:90276413
 A:Accession: S10572
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
 A:Cross-references: EMBL:X52229; NID:g37053
 R:Wreschner, D.H.
 submitted to the EMBL Data Library, March 1990
 A:Reference number: S40293
 A:Accession: S40293
 A:Molecule type: mRNA
 A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
 A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054
 R:Abe, M.; Siddiqui, J.; Kufe, D.
 Biochem. Biophys. Res. Commun. 165, 644-649, 1989
 A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associated
 A:Reference number: A36735; MUID:90088473
 A:Accession: A36735
 A:Molecule type: mRNA
 A:Residues: 1-142,'Q',144-162,'Q',164-168 <ABE>
 A:Cross-references: EMBL:M31823; NID:g181542; PIDN:AAA35757.1; PID:g181543
 R:Masuzawa, Y.; Miyauchi, T.; Hamaouque, M.; Ando, S.; Yoshida, J.; Takao, S.; Shimazu, H
 J. Biochem. 112, 609-615, 1992
 A:Title: A novel core protein as well as polymorphic epithelial mucin carry peanut agglu
 A:Reference number: JX0235; MUID:93123189
 A:Accession: PX0066
 A:Molecule type: mRNA
 A:Residues: 998-1011,'ES',1014-1017,1018-1032,'T',1034-1037,1038-1057 <MAS>
 A:Experimental source: gastric carcinoma cell
 R:Erhan-Licht, S.; Baruch, A.; Elroy-Stein, O.; Keydar, I.; Wreschner, D.H.
 FEBS Lett. 356, 130-136, 1994
 A:Title: Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins cytokine
 A:Reference number: S51026; MUID:95080414
 A:Accession: S51026
 A:Contents: annotation
 A:Note: undetermined tyrosine residues in the carboxyl-terminal non-repetitive region ar
 C:Comment: This protein is length polymorphic. Individuals may have between 21 and 125 c
 partial repeats. The repeat shown is defined by SmaI nuclease sites.
 C:Comment: Serine and threonine residues in the tandem repeat domain are extensively gly
 C:Comment: For an alternative splice form without a tandem repeat domain, see PIR:S48146
 C:Genetics:
 A:Gene: GDB:MUC1; PUM
 A:Cross-references: GDB:120705; OMIM:158340
 A:Map position: 1q21-1q23
 A:Introns: 20/1; 62/3; 1165/3; 1184/2; 1230/1; 1270/3; 1320/3
 C:Superfamily: polymorphic epithelial mucin
 C:Keywords: alternative splicing; duplication; glycoprotein; phosphoprotein; polymorphis
 F:1-1344/Product: mucin 1 precursor, splice form A #status predicted <PREA>
 F:1-62/Region: mucin 1 amino-terminal non-repetitive

F:1-23/Domain: signal sequence #link PREA #status predicted <STGA>
 F:1-19,29-32/Domain: signal sequence #link PREB #status predicted <SIGB>
 F:1-19,29-1344/Product: mucin 1 precursor, splice form B #status predicted <PREB>
 F:1-19,29-212,1033-1344/Product: mucin 1 precursor, epithelial tumor antigen splice f
 F:138-1017/Region: 20-residue repeats (GSTAPAHGVTSAPDTRAP)
 F:1143-1344/Region: mucin 1 carboxyl-terminal non-repetitive
 F:1245-1272/Domain: transmembrane #status predicted <TRM>
 F:1046,1064,1118,1144,1222/Binding site: carbohydrate (Asn) (covalent) #status predic
 F:1213/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 10.2%; Score 672; DB 1; Length 1344;
 Best Local Similarity 27.8%; Pred. No. 7.6e-23;
 Matches 310; Conservative 78; Mismatches 472; Indels 256; Gaps 56;
 QY 101 LTYNKETTVEKTTTTNK--QTSTGKETTSAKETQSTIEKTSKD-LAPTSTKVLAAPT 157
 DB 15 LTVLTATTAPKATVVTGSHASSTPGGKETATSQRSSVPSSTKNAVSMTSSVLSSHS 74
 QY 158 P-KAETTTKGP--ALTTKPEP-----TPTTKPEPASTTP-----KE 190
 DB 75 PGSGSSSTTQGDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNK 134
 QY 191 PPTTT-----TKSAPTTKPEAPTTT-----KSAPTTKPEAPTTTKEP-----A 230
 DB 135 PAPGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTS 190
 QY 231 PTTTPKEPAPTTTKEPAPTTTTSAPTTKPEAPTTT-----KKPAPTTTKEP----- 276
 DB 191 PDR--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGHGVTSAPDT--RPAGSTAPPAGH 247
 QY 277 --APTTTKEPPTTP-----KEPAPTTKEP-----APTTPKEPAPTAPK----- 313
 DB 248 TSAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPPAGH 307
 QY 314 KPAPTTPKEPAPTTTPK-----EPAPTTTKEPSTTPKPEAPTT--TKSAPTTTKEPAPTT 366
 DB 308 TSAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPPAGH 367
 QY 367 TKSAPTTTKEPSTTPKPEP-----APTTPKEPAPTTPK-----KPAPTTPKPEPAPTTPK- 415
 DB 368 T-SAPDT--RPAGSTAPPAGHGVTSAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPP 424
 QY 416 ---EPAPTTTTPKAPITAPKPEAPTTTKEPAPTTPKKTPTTPEKLAAPTTPKEPAPTTP- 471
 DB 425 HGVTAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPP 484
 QY 472 ---ELAPTTPEEPTTP-----EEPAP--TTPKA-----AAPNTPKEPAPTTPK- 511
 DB 485 HGVTAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPP 544
 QY 512 ---EPAPTTTKEPAPTTTKEPAPTTPKGTAPTTTKEPAPTTP-----KKPAPKEL 558
 DB 545 HGVTAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGST 600
 QY 559 APTTTKEPSTTSDKPAP--TTPKGTATPTTKEPAPTTPKPEAPTTPKG----TAPTTLK 612
 DB 601 AP-PAHGVTSAPDTRPAPGSTAPPAGHGVTS-----APDTRPAPGSTAPPAGHGVTSAPDTR 655
 QY 613 EPAPTTT-----KKPAPKELAPTTTKEPSTTSDKPAP--TTPK-----ETAPTT 656
 DB 656 APGSTAPPAGHGVTSAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPPAGHGVTSAPDTR 714
 QY 657 KEAPTTTP-----KKPAPTTPEPTTSEVSTPTTKEPPTT-----IHKSPD 699
 DB 715 PAPGSTAPPAGHGVTSAPDTRPAPGS--TAPPAHGVTSAPDTRPAPGSTAPPAGHGVTSAPD 772
 QY 700 ESTPELSAETPKALENSPKFEGVPTTKTAA-----TKPEMTTAKDK-----TTERD 748
 DB 773 TRPAGSTAPPAGHGVTSAPDTRPAPGSTAPPAGHGVTSAPDTRPAPGSTAPPAGHGVTSAPD 832
 QY 749 LRTTPEIT-----TAAPKMTKETATTTETKTSKITATTQVTTSTTQDTP--EKIT 799

Db	833	TRPAGSTAPPAGHVSAP-----DTRPAGSTAPPAGHVSAPDTRPAGSTAPPAGHVT	888
Qy	800	TLKTTTLAPKVTITTKTITTT--EIMNKPEETAKPKDRATNSKATTPKPKTKAPKKT	857
Db	889	SAPDTRPAGSTAPPAGHVSAPDTRPAGSTAPPAGHVSAPDTRPAP--GSTAPPAHG	946
Qy	858	STKKPTMPRVKPKTTPPKMTSTMPELN-----TSRIAEAMLOT--TTRP---NOTPN	909
Db	947	VTSAPDTRP---APGSTAPPAGHVSAPDTRPAGSTAPPAGHVSAPDTRPAGSTAPP	1003
Qy	910	SKLVEVNPXSEDAGGAGETPHMLLRPHVEMPEVTDMDYLRPVNNOGIIINPMLSDETN	969
Db	1004	AHGVSAPDTRPAGSTAPPAGH-----GVTSAPDNRNALGSTA-----PPVHNVT	1049
Qy	970	ICNGKPVDTLLTNRGTLVAFRGHYFWMLSPFSPS	1005
Db	1050	ASGSASGSTLVHNGTSARATTPASKSTPFSIPS	1085
RESULT 9			
Tl8535			
high molecular mass nuclear antigen - chicken (fragment)			
C:Species: Gallus gallus (chicken)			
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999			
R;Shimada, K.; Harata, M.; Mizuno, S.			
J. Cell Sci. 110, 3031-3041, 1997			
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick			
A:Reference number: Z18955; MUID:9803440			
A:Accession: Tl8535			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: mRNA			
A:Residues: 1-1151 <SHI>			
A:Cross-references: EMBL:D88440; NID:d117138; PID:d1025045; PIDN:BAA24137.1			
Query Match 10.2%; Score 671; DB 2; Length 1151;			
Best Local Similarity 25.9%; Pred. No. 7.2e-23;			
Matches 284; Conservative 100; Mismatches 419; Indels 292; Gaps 55;			
Qy	38	TPKP-----PVVDEAGSLDNGDKVITPDTSTTQHNVSTIS--PKIITAKPINRPS	88
Db	113	TPPSQGPAGTPPPSQAAGAPKGDGAQPSGTSKADGKPAADVPKAITA--ATEARP--	170
Qy	89	LPPNSDTSKETSLTVNKETTVETKETTNN--OTISDGRKETSASAKTOSIEKTSADKL	146
Db	171	-----ASAASPTVPKATAEATAVTAASQAPKAAATAAAVTA--SQSAPKATV-EV	219
Qy	147	APTSKVLAKPTPKAETTTKGPAITPKPTP--TTPKEPASTTPKEP-----TPPTIKSAPT	201
Db	220	KPAAAVAKAEKAVTAAAAAPKATAEAKPAPVTSPTIPCSAAEKPLTAASPTASKA--T	277
Qy	202	TPKEPAPT-----TKSAPTTPKEPAPTTK-----	227
Db	278	ABAKVPVPATASIMATKVTAEAKPAPSPVPRKATIDIKAVTATAPKAGDPVPAVAVCAEA	337
Qy	228	EPAPTTPKEPAPTTPKEPAPTTPKSAPT-----PKEPA-----PTTPKPKPAPTTPK	274
Db	338	KFAPPPPPQQLPKAAAAAAPTGLKPKATAPPHGSFRANSHVTVTTPNVPRAAAAATVP-	396
Qy	275	EPAPTTPKEPPTT-----PKEPAPTTPKEPAPTTPKEPAPTAPKPKPAPTTPKEPAPTTPKE	330
Db	397	-TAGAVPKRASTGTTTAAAFQOPVP---KAAPVTPSPQOAVPRAATAAA-----APVTPQQ	448
Qy	331	P---APTTPKESPTTPKEPAPTTPKSAPTTTPKEPAP-----TTTKSAPTTPKEPSPTTTKE	384
Db	449	PVTKAATTNATPPQPIPKAATTTATVTPQQIPKAGTDAAPPVAPVAPRAPSNGRAAT	508
Qy	385	P-----APTTPKEPAPTTPKKPAPTTPKEPAP-----TTPKEPAPT--TTKKPAPTAP	430
Db	509	PGVPAATDPKPPPTPQSVPSAVTEPKPQPRAAAPPSPNEATPAVPSPLKSPLEPTIP	568
Qy	431	K-----EPAPTTPK---ETAPTTPKLTPTTPEKLAPTTPKEPAPTTPPELA-----	474

Db	569	KPVPLMALTPOPVTAQMVTLQAATKPSPIVPKASPK-ALMTPPPPPPGPRALAAAKLLG	627
Qy	475	-PTTP-----EETPTTP---EETPTTP-----KAAAPNTPK	504
Db	628	LPSSSVASAMHAKVTPRPLPASVPVMAASPASGLDGAARVALATINAAASFGAKPEAAGONG	687
Qy	505	-----PAPTTPKEPAPTTPKEPAPTTPKETAPT-----TPKGTAPT-----	540
Db	688	TLMAPGAAANTQMTAIGRAAGAAQTAPMGAAHTHVSPMGAGGATQMSPTGAANTHMSPIGA	747
Qy	541	-----TLKEPAPTTPKPKAPKELAPTTTKEP-----TSTSDKPAPTTP	579
Db	748	GCATQMSPMGAANTQMSPMGAATTTQMSPMGAAATTPSPMGAAATQVTSAGNTMQVSP	807
Qy	580	KG--TAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEP--APTTPKKPAPKELAPTTTKEP	635
Db	808	MCAATPPQTPSVGAATTP--QPSFM-----CAATTLMSPMCAATTPQ---PSPMGAVTTPP	859
Qy	636	-----TSTTSKDP--APTTPKET---APTTPKEP---APTTPKKP--APTTPETPP-----	676
Db	860	PMAATNTTQPPMAASTPQSTPMGAATTTQSPMGATTTQSPMGASTPQAPTTPVAGSPT	919
Qy	677	PTTSEVSTPTTTPKEPTTIHKSDESTEPEL--SAEPTPKALENSPKPEGVPTTKTTPAATKP	734
Db	920	PPPTTPPSPTAQTSPQPMSPKSPPPDPKAPSAQAQTSAAHVANASPGV--TAVSPA---P	975
Qy	735	EMTTAKDKTTERDLRTTTPETTTAAPTMMT--KETATTTKTTESKITATTTQVTSSTTQDT	793
Db	976	IGVTEASPSADGARLSFGPTAATDGPASPAATADVTEAATD--VTAAATAVPA-----EA	1029
Qy	794	TPFKITTLKTTTLAPKVTITTKTITTTTEIMNKPEETAKPKDRATNSKATTPKPKTKAP	853
Db	1030	AP-----TKAKRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSPASPA---	1077
Qy	854	KKFTSTKKPKTMRVRKPTTTPRKMTSTMPELNPTSRIAEAMLOTTRPRNOTPNSKLV	913
Db	1078	-----PAVGDGGOOMTPGAQSVPP-----VTEAAVQ-----	1104
Qy	914	FVNPKSEADAGGARGE	928
Db	1105	EAAAAAAGAE	1119
RESULT 10			
T25697			
hypoetical protein F16F9.2 - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000			
C:Accession: T25697			
R;Fulton, B.			
submitted to the EMBL Data Library, August 1996			
A:Description: The sequence of C. elegans cosmid F16F9.			
A:Reference number: Z20071			
A:Accession: T25697			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-1229 <FUL>			
A:Cross-references: EMBL:U67956; PIDN:AAB07691.1; GSPDB:GN00028; CESP:F16F9.2			
A:Experimental source: strain Bristol N2; clone F16F9			
C:Genetics:			
A:Gene: CESP:F16F9.2			
A:Map position: X			
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3			
Query Match 10.0%; Score 659.5; DB 2; Length 1229;			
Best Local Similarity 29.0%; Pred. No. 2.5e-22;			
Matches 307; Conservative 78; Mismatches 339; Indels 333; Gaps 52;			
Qy	57	KVTT--PDTSTTQ---HNKVSTSPKTTTA-KPINRPSLPNNSDT-----SKETSILVN	104
Db	119	KVTTSTDASTNAPTGTGKDTTPEITGIVINSKESVTDMSVTRSTTLLSPTELLTS	178

	conservative	33.	Mismatches	252:	Indels	44;	Gaps	7;
--	--------------	-----	------------	------	--------	-----	------	----

[illegible]

RESULT 13
C84672
hypothetical protein At2g27380 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
A:Accession: C84672
R:Lin., X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujita,
T.; Lin, A.C.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.;
Moo, H.; Kooh, B.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.;
Eickbusch, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.;
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
A:Reference number: A84420; MUID:20083487
A:Accession: C84672
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-761 <STO>
A:Cross-references: GB:AEO02093; NID:g5306260; PIDN:AAD41992.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g27380
A:Map position: 2

Query Match	9.6%	Score 632;	DB 2;	Length 761;	
Best Local Similarity	31.1%	Pred. No. 2.5e-21;			
Matches 217;	Conservative	44;	Mismatches 354;	Indels 82;	Gaps 33

QY	156	PTPKAETTITKGPALTPT	---	KEPTPT	-TPKEPASTTPKEPTPT	TTIKSA	PTPTTK	KEPAPT	TT	211
			:		:		:		:	
Db	69	PPPIQKPPIYSPPIYPPIQKPEPT	PTTYS	SPPIYPPIQKPEPT	PTTYS	SPPIYPPIQKPEPT	PTTYS	SPPIYPPIQKPEPT	PTTYS	128
			:		:		:		:	
QY	212	KSAPTPTTKEPAPT	TTKEPAP	---	TTTKEPAPT	TTTKEPAPT	TTTKEPAPT	TTTKEPAPT	TTTKEPAPT	263
			:		:		:		:	
Db	129	YSPIYP	---	PPPIQKPEPT	PSYSPVVKPPPV	---	QMPPTT	YSPIYPPIQKPEPT	PTTYS	181
			:		:		:		:	
QY	264	TPKKP	---	APTTPKEPAPT	PTT	---	KEPTPT	---	TTTKEPAPT	313
			:		:		:		:	
Db	182	PPIKPPVHKPEPT	PTTYS	PPIKPPVHKPEPT	PTTYS	PPIKPPVHKPEPT	PTTYS	PPIKPPVHKPEPT	PTTYS	241
			:		:		:		:	

512 EPA-----PTTPKPEAPTTPKETAPTTPKGTAPTTLKEP 545
 6572 EVAFEENVTHVEEYLVEEEEEEIHHEEEFITEEEVVPVLPVR-VPEVPRKKVPDEE-KKP 6629
 546 APTTPKKPAPKELAPTTTKEPTSTT-----SDKPARTPKGTAPTTPKBPAPTTPKPA 599
 6630 VPVPPKKEAPPAPKVPVPPKPEEKVPVLPKKEP---PPAKVPEVPKKPVP-EEEKVPV 6684
 600 PTPPKGTAPTTLKEPA--PTTPKKPAPKEL-----APTTTKGPTSTTSDDKPARTPKETAP 653
 6685 VPVKKVEAP-----PAKVPEVPKKPVPKVPAPKVPKVEAPPAPKVPVPEVPPKKLIPKKEKP 6739
 654 T-TPKE---PAPTTPKKPAPTTEPTDPPTTSEYSTPTTTTKEPTTIHKSPDESTPELSAEP 709
 6740 TPVKKVEAPPKVPKPKREPPVPVVALPQEEVLFEEIYPVEEVLPDEEEVLPDEE-EV 6798
 710 TPRALENSPKPEGVPTTKTPAATKPEMTTTAKOKITTERDL--RTTPTTTAAAPKMTTETA 767
 6799 LPDEEEVLPDEEEIPEEVEEVPVEEVEEVPVEEVEEVLPEVKVPVVPAP-----VP 6853
 768 TTEKTTTESKITATTQVTSITTTQDTPPKITTLTKTTTLAPKVTTKKTIITTTIMNKPE 827
 6854 EIRKKVTEKKVVPKKEEAPPAPKVPVPPKVE--EKRIILPK-----EEVLPVEVTEBEPE 6907
 828 ETAPKPKDRATNSKATTPKPKOPTKAPKPTSTKK-----PKTMPRVRK---PKT-TPTPR 878
 6908 E-----EPISSEETPREPPSEIEEVEEAPPVPEVIRKKVPEAPTTPVKP 6951
 879 KMTSTMDELNPTSRIAEAMLOTTRPNQTPNSKLVENPKSEDAGGAGSETPHMLLRPHV 938
 6952 KVEA--PPAKVSKKILPEEKVPVPPVQKKAPPAKVPVE-----PPAKGRTVLEEKVSAFR 7021
 939 FMPEVTPDMYDLPRVPNGOGLIINPMLSDETNICNGKPVVDGLTTLRNGTLVAFR 991
 6987 --PKVPEKKVL--VPKKEAV-----PPAKGRTVLEEKVSAFR 7021

RESULT 12

Extensin class 1 precursor - cowpea
 Species: Vigna unguiculata (cowpea)
 Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 17-Nov-2000
 Accession: T11622; S54155
 Author: Arsenijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
 Mol. Plant Microbe Interact. 10: 95-101, 1997
 Title: Rhizobia modulate root-hair-specific expression of extensin genes.
 Reference number: Z17301; MUID: 97155574
 Accession: T11622
 Status: preliminary; translated from GB/EMBL/DBJ
 Molecule type: DNA
 Residues: 1-489 <ARS>
 Cross-references: EMBL:X91836; NID:g1015936; PIDN:CAA62943.1; PID:g1015937
 Experimental source: sub-species Red calcoana
 Author: Arsenijevic-Maksimovic, I.; Broughton, W.J.; Krause, A.
 Submitted to the EMBL Data Library, April 1995
 Description: A class of root-hair specific extensins involved in rhizobium/legume inte
 Reference number: S54155
 Accession: S54155
 Status: preliminary
 Molecule type: mRNA
 Residues: 326-489 <AR2>
 Cross-references: EMBL:X86030; NID:g791149; PID:g791150
 Genetics:
 Gene: Ext26G
 Superfamily: hydroxyproline-rich glycoprotein
 Keywords: glycoprotein; hydroxyproline
 F;1-23/Domain: signal sequence #status predicted <SIG>
 F;24-489/Product: extensin class 1 #status predicted <MAN>

Query Match 9.6%; Score 633; DB 2; Length 489;
 Best local similarity 31.2%; Pred. No. 1.5e-21;

QY	314	KPAPT--TPKEAPRTTPKEAP--	---TTTKESPRTTPKEAPRTTTKSPARTTTKEAPRTTTSK	369
Db	242	PRTPIYSPPIKPPPVHKPRTPIYSPPVKPPVQVQRTPTPIYSPVHKPRTPTYSPP	301	
QY	370	ABTTPEKESPRTTTPKEP--	---ABTTPEKAPRT--TPKKAPRTTPKEAPRTTPKEAPRTTTK	424
Db	302	VKSPVPVQKPPRTTYSPPRIKPPPVQKPPRTPIYSPPIKPPPVKPPRTPIYSPPVXPP--PPVHKP	360	
QY	425	PAP---TAPKEAPRT--	---TPKETAPRTP---KKLTPRTPEKLPARTTPKEKPARTT---PE	471
Db	361	PTPIYSPPVKPPPVHKPRTPIYSPPVKPPPIQKPPRTTYSPPRIKPPPLQKPPRTPIYSPPI	420	
QY	472	ELAPRTTPBERPTTPPEAPRT--	---TEKAAAP--NTPEKAPRTTPKEAPRTTPKEAPRTTP	526
Db	421	KLPPVKKPRTPIYSPPVKPPPVHKPRTPIYSPPVKPPPVHKPRTTYSPPRIKPPPVKPPRT	480	
QY	527	KETAPRTTPKGARTTTLKEAPRT--	TPKKPAPKELAPRTTTTKEPTSTTSKPARTTPKGTAP	584
Db	481	TYSPVPQVQVQ---PPVQKPPRTTYSPPVKKPPPIQKPPRT--	---PTYSPPIKPPPVKPPRTTPT	532
QY	585	TPKEAPRTTPKEAPRT--TPKGARTTTLKEAPRT--	TPKKPAPKELAPRTTTKGTSTTSD	641
Db	533	SPPIKP--PPVHKPRTTYSPPRIKPPPVHKPRTTYSPPRIKPPPVHKPRTTYSPPIKPPPP	591	
QY	642	KPAPTTPKETAPRTTPKEAPRTTPKKKAPRT--	TPETPPPTTSEVSTTPTTK--EPRTIHKSP	698
Db	592	VHKPRTTYSPPRIKPP--	---PPVHKPRTTYSPPRIKPPPVKPPRTTYSPPRIKPPPVHKP	647
QY	699	DESTPELSAEPPTKALENSKPEGVRTTTPAATKPEMTTAKDKTTEDRLTPTPETTTA	758	
Db	648	---TPRTYSPPRIKPPPVQKPPRTTYSPPVAPPPVQVPP--	TPRTYSPPVKKPPPVQVPPRTTYS	703
QY	759	APKMTKETAATTTKETSKITATTTQVSTTTQDTP	795	
Db	704	PPVKKPPPVQVPPRTTYSPPRIKPPPVQVPPRTTPTSP	740	

RESULT 14
T34513
hypothetical protein zk783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favelllo, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid zk783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:zk783
A:Experimental source: strain Bristol N2; clone zk783
C:Genetics:
A:Gene: CESP:zk783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/3
3504/1

RESULT 15

RESULT 13
 151618
 nucleolar phosphoprotein - African clawed frog
 C:Species: xenopus laevis (African clawed frog)
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996
 C:Accession: U51618; S57757
 R:Cairns, C.; McStay B
 J. Cell Sci. 108, 3339-3347, 1995
 A:Title: Identification and cDNA cloning of a X

Mon Apr 29 08:35:14 2002

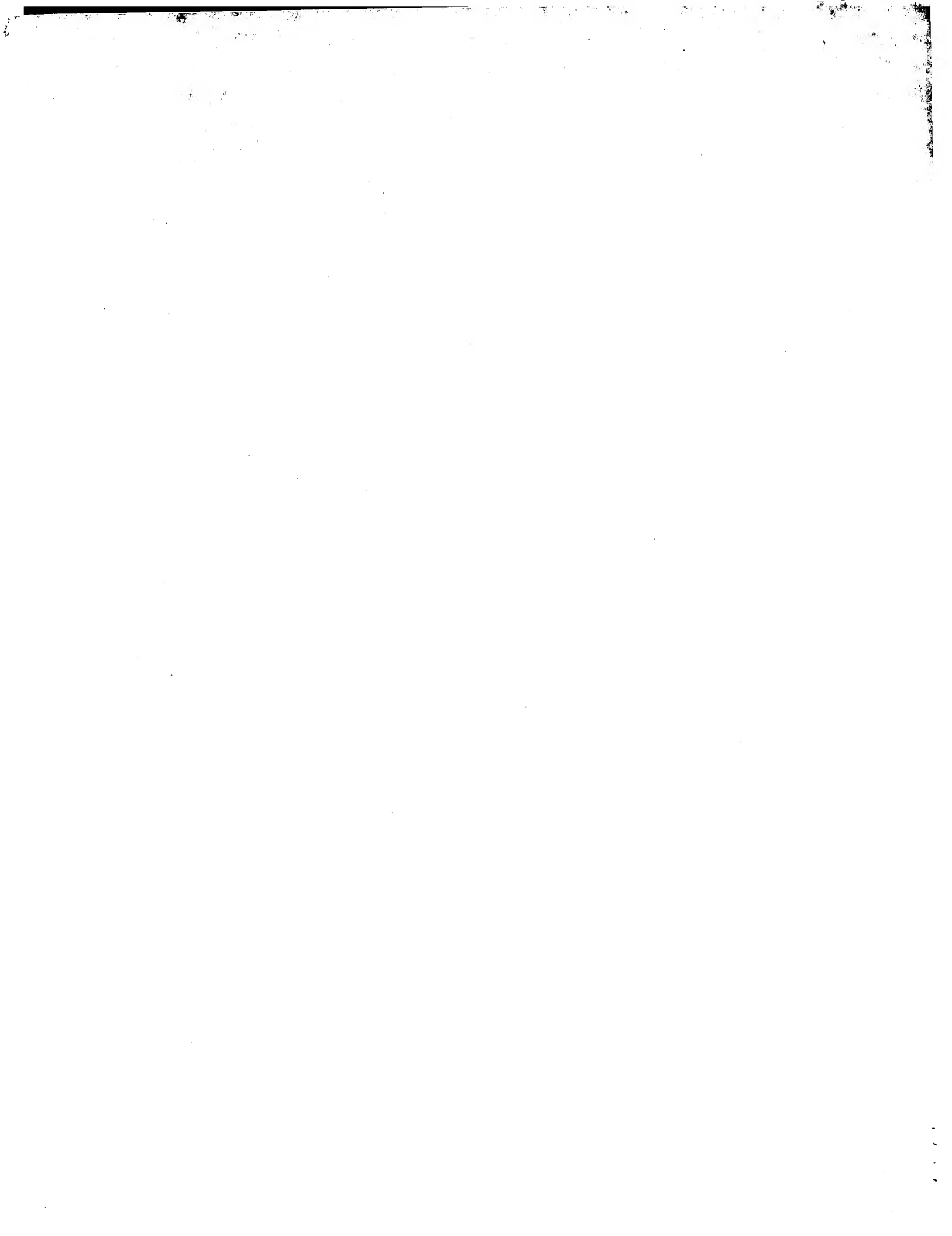
A:Reference number: I51618; MUID:96019267
A:Accession: I51618
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-990 <CAI>
A:Cross-references: EMBL:X88927; NID:9895920; PIDN:CAA61368.1; PID:9895921
C:Genetics:
C:Gene: xNop180
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

Query Match 9.5%; Score 627; DB 2; Length 990;
Best Local Similarity 26.2%; Pred. No. 5.5e-21;
Matches 258; Conservative 134; Mismatches 401; Indels 192; Gaps 46;

QY 7 PIYLLLLSVFIQVSSVKDNKRNKTKKPTPKPPVVDGAGSLDNGDFKVTTPDTSTT 66
DB 44 PLSLLDIFSDV-----KSPDAKKRPPANGLPKKSAKESSEDSEDEPPAKKRA 98
QY 67 QHNKVTSPKITTAKPINRPSPNPSDTSKETSITVKNKETTVE--TKETTTNK-----Q 120
DB 99 Q-----PAGGKKPVVQVKKAKSSSEDSDSD-----SEETKKPPAKRPAQTPKVAAYK 151
QY 121 TSTDGKEKTTSAKETOSIEKTSKOLAPTSKVLAKPTPKAETTTKGPALTTKEPTPTTP 180
DB 152 TPTQKRAKSSSS-ESSSSEDEASKKQPVIVK-----PPKQAVVYKAGLANNNGKTADSSSS 206
QY 181 KE-----PASTTPKEPTPTTILKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTTPK 236
DB 207 EDSDDPPAKKTAATKTPPT--KPATAAKPQAKKTAGKKSSREDSDSDDEQKTAASK 263
QY 237 PAPTT-TKEPAPT-----TKSAPTTKPEAPT-----TPKKPAPTTPKEAPTT 280
DB 264 PKPDVSAVPPPTSVSKKKTLSQGTAKPESDSDSDSDDEEQPAKKIVPAKAAASA 323
QY 281 P-----KEPTPTTKEPAPTTPKEAPTTKPEAPTAPKPKPAPTTPKPEAPTT---KEPA 332
DB 324 PKLAKKAETSTDSESDSDSDDEKSSKVLGVKAAAPK-APAAP--DAKSTPVAAAKKSA 380
QY 333 P-----TTKPESTTPKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKSAPTTPK 375
DB 381 PAKKASSSDSDSDSDNEETTTKPAAKTTPAKSAATPTSKTPTNGK---ATPTSKTAPKPG 437
QY 376 EPSPTTTKEPAPTTPKEPA---PTTPKKPAPTTPKPEAPT--TPKEAPTTTKKPPAPTAP 430
DB 438 TPKTSTAKKSDSDSDSDDEETTTKPAKTTTPAKSAATPTSKTP--TNSKATPTSK 494
QY 431 KEPA-PTTPKETA-----PTTPKKLTPTPEKLIAPTTPEKAPTTPEELAPTT 477
DB 495 KTPAKPGTPKTSAAKKDSSSDSDSDSDSEKKTPEA--KRAAKTTPAKPA-----AKTT 545
QY 478 PEEPTPTTPEPAPTTP-KAAANTPKPEAPT---TPKEAPTTTPKEAPTTTPKETAPT 532
DB 546 PAKPA-----AKTTPAKPAKSTPGQVPTKKESSDSDSDSDDEKSSAKPAVK 598
QY 533 TP-KCTAPTTLKEPAPTTPKPAKELAPTTTKETPTSTSDKPAPTTPKGTAPTTPKEPA 591
DB 599 TPGKATS-----KPVVASKPVPAK-----KASSSDSDSDSEETTKTKPLTKLSPA 645
QY 592 PTT--PKEPAPTTPKGTAPTLKEPAPTTPKPKAPKELAPTTTKP-----TSTSD 641
DB 646 VKTLPPKKAESSDSDSDSEK---TKPAKPPAKSAIPVNTKAPQONKASKASCSDSD 702
QY 642 KPAPTTKETAPT--TPKEAPTTPKK-PAPTTPTPTPTTSEVSTPTTTKE-PTTIHKS 697
DB 703 SSSEEGKSKQPTGKSPAATAKATAPKKNPVAVNKDPSSSSDSDSDSGDDEKQKQAAAA 762
QY 698 PDESTPELSAETPKALENSKPEPGVPTTKTPAATKPEMTTAKDKTTTERDLRTTPTTT 757
DB 763 KDVKOGAKAAKFTPKKAASSSE---DSSDSDSDSVKAKKTNFAVSKSPV---TPKAVP 815
QY 758 AAPKMTKETATTEKTESKITATTTQVTSITTTQDTTPPKITTLKTTTLAPKVTTK--K 815

DB 816 AAKESSESSESDSEDEKQGGKNTSTTKIANST-----PKAAAAECSE 857
QY 816 TITTTTEIMNKPEETA-KPKDRATNS---KATTPKPOKPTKAPKPKTSTKPKTMPVRKP 871
DB 858 ESSSSEDEGRANGTSGRRKRESTGNACEAVTPE-----NKKLKAKSPNTFFPKVKK 909
QY 872 KTTPTP-RKMTSTMPELNPTSRIAE 895
DB 910 ELKNTPPRRVVEEDIEINP--RMAD 932

Search completed: April 26, 2002, 16:30:58
Job time: 653 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:22:50 ; Search time 62.75 seconds
(without alignments)
718.105 Million cell updates/sec

Title: AAL
Perfect score: 6568
Sequence: 1 MAWKTLPIVLLLSVEVIQ.....ARAITRSGTLLSKVWYNCP 1229

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1168	17.8	5179	MUC2_HUMAN	Q02817 homo sapien
2	950	14.5	1664	SLPL_CLOTM	Q06852 clostridium
3	789.5	12.0	1367	AMYH_YEAST	P08640 saccharomyc
4	651	9.9	1235	MUC1_HUMAN	P15941 h mucin 1 p
5	556.5	8.5	875	FPI_MYTED	Q25460 mytilus edu
6	555.5	8.5	2700	ZAN_HUMAN	Q9v493 homo sapien
7	551	8.4	620	EXTN_TOBAC	P13983 nicotiana t
8	533	8.1	1087	NFH_MOUSE	P19246 mus musculu
9	530.5	8.1	1162	TCNA_TRYCR	P23253 trypanosoma
10	530	8.1	865	CPN_DROME	Q02910 drosophila
11	519	7.9	872	FPI_MYTGO	Q25434 mytilus cor
12	509.5	7.8	662	MUC1_XENLA	Q05049 xenopus lae
13	503.5	7.7	1970	RPB1_HUMAN	P24928 homo sapien
14	498.5	7.6	1970	RPB1_MOUSE	P08775 mus musculu
15	497	7.6	831	NFH_RAT	P16884 rattus norv
16	493.5	7.5	467	RPB1_CRIGR	P11414 cricetus
17	488.5	7.4	826	SNP2_PLAYO	G01443 plasmodium
18	488	7.4	1020	SSP2_PLAYO	P12036 homo sapien
19	475.5	7.2	267	EXTN_MAIZE	P14918 zea mays (m
20	471.5	7.2	5376	ZAN_MOUSE	O88759 mus musculu
21	468.5	7.1	634	HWPI_CANAL	P46593 candida alb
22	467	7.1	817	VRP1_YEAST	P37370 saccharomyc
23	463	7.0	2142	BAT2_HUMAN	P48634 homo sapien
24	459	7.0	1161	YJ9P_YEAST	P47179 saccharomyc
25	454	6.9	797	VGLX_HSVB	P28968 equine herp
26	448.5	6.8	670	VG50_HSV11	Q00130 ictalurid h
27	442.5	6.7	751	FPI_MVTGA	Q27409 mytilus gal
28	439.5	6.7	1083	T2D3_HUMAN	O00268 homo sapien
29	432.5	6.6	439	XP2_XENLA	P17437 xenopus lae
30	432	6.6	3164	TEGU_HSV11	P10220 herpes simp
31	426.5	6.5	1794	YAV1_SCHPO	Q10172 schizosacch
32	426	6.5	2715	TRX2_HUMAN	Q9unn6 homo sapien
33	424.5	6.5	1125	MAP4_MOUSE	P27546 mus musculu

34	424	6.5	307	1	SGS3_DROME	P02840 drosophila
35	424	6.5	1251	1	YQUS_CAEEL	Q09550 caenorhabdi
36	422.5	6.4	3421	1	TEGU_HSVB	P28955 equine herp
37	421	6.4	1229	1	NI21_HUMAN	Q9v2n3 homo sapien
38	419.5	6.4	2476	1	ZAN_PIG	Q28983 sus scrofa
39	419.5	6.4	2774	1	MAP4_RAT	P34926 rattus norv
40	419	6.4	1185	1	DRPL_HUMAN	P34259 homo sapien
41	411.5	6.3	907	1	VGP3_EBV	P03200 epstein-bar
42	410	6.2	1183	1	DRPL_RAT	P54258 rattus norv
43	410	6.2	1411	1	TCOF_HUMAN	Q13428 homo sapien
44	407.5	6.2	3256	1	KI67_HUMAN	P46013 homo sapien
45	405.5	6.2	2517	1	NCR2_HUMAN	Q9v618 h nuclear r

ALIGNMENTS

RESULT	1
MUC2_HUMAN	
ID	MUC2_HUMAN
AC	Q02817; Q14878; STANDARD; PRT; 5179 AA.
DT	01-JUN-1994 (Rel. 29, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
GN	MUC2 OR SMUC.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Intestine;
RX	MEDLINE=94132002; PubMed=8300571;
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT	"Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT	Identification of the amino terminus and overall sequence similarity
RT	to prepro-von Willebrand factor.";
RL	J. Biol. Chem. 269:2440-2446(1994).
RN	[2]
RP	SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC	TISSUE=Colon;
RX	MEDLINE=93016075; PubMed=1400449;
RA	Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA	Kim Y.S.;
RT	"The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT	both upstream and downstream of its central repetitive region.";
RL	J. Biol. Chem. 267:21375-21383(1992).
RN	[3]
RP	SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX	MEDLINE=91358717; PubMed=1885763;
RA	Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA	Petersen G.M., Kim Y.S.;
RT	"MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT	and polymorphism.";
RL	J. Clin. Invest. 88:1005-1013(1991).
CC	-I- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC	OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC	PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC	AGENTS AT MUCOSAL SURFACES.
CC	-I- SUBUNIT: MULTIMERIC.
CC	-I- SUBCELLULAR LOCATION: SECRETED.
CC	-I- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC	BRONCHUS, CERVIX AND GALL BLADDER.
CC	-I- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC	INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC	-I- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC	VARIES AMONG DIFFERENT ALLELES.
CC	-I- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC	OF SILKWORM HEMOCYTIN.
CC	-I- SIMILARITY: CONTAINS 2 VWFC DOMAINS.
CC	-I- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).
CC	-----

QY 881 -----TSTMPDLNPTSRIRAMLOTTTRPQNTNSKLVNPKSEDAGAGETPHM- 932
 Db 1454 YLRGYPDGSFRPNRITRAEAVIF-----AKLL-----GADSYGAQASPSYSD 1498
 QY 933 LLRPH-----VFME-----VTPDMYLPVPPNOGI----- 958
 Db 1499 LADTHAANAWEKATSGLEKGYGPDGTFKPDQNTIRAEFATVVLHFLTKVKGQEIEMSKLA 1558
 QY 959 ---IINPLMSDETINIC-----KPDVGLTTL-----RNGTLVAFRGHYFWMLSPSPSPS 1005
 Db 1559 TIDISNPKFDD-----CVGHWAAQEFIEKLSLGYISYIPDGT-----FKPON 1600
 QY 1006 PARRITEVWGPISPIDTVFRNCCEKCTFFFKD--SQYWRFTNDIKD 1050
 Db 1601 YIKRSEV-----ALINRALERGLNCAKPLFPDVNENYNAF-CDIMD 1642
 RESULT 3
 AMVH_YEAST STANDARD; PRT; 1367 AA.
 AC P08640; P08068;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GLUCOAMYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-
 DE GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).
 GN STAL OR STA2 OR MAL5 OR VIR019C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID:4932;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Church C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 RA Gaitanaris G., Hamlyn N., Harsnall T.S., Hunt S., Jagels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moulton N., Odeil C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC 1994) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
 RP MEDLINE-87194600; PubMed-3106330;
 RX Yamashita I., Nakamura M., Fukui S.;
 RT "Gene fusion is a possible mechanism underlying the evolution of
 RT STAL";
 RL J. Bacteriol. 169:2142-2149(1987).
 [3]
 RN SEQUENCE OF 1-31 FROM N.A.
 RP STRAIN-SPX101-1C;
 RX MEDLINE-89031230; PubMed-3141213;
 RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
 RT "Similar short elements in the 5' regions of the STA2 and SGA genes
 RT from Saccharomyces cerevisiae";
 RL FEBS Lett. 239:179-184(1988).
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
 CC GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
 CC WITH RELEASE OF BETA-D-GLUCOSE.
 CC -!- SIMILARITY: TO S-POMBE SPBC215.13.
 CC -!- SIMILARITY: SOME, TO S-POMBE SPCC285.13C.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z38061; CAA86176.1; -;
 CC EMBL; M16164; AAA35014.1; -;

DR EMBL; M16165; AAA35015.1; -;
 DR EMBL; X13857; CAA32069.1; -;
 DR PIR; B26877; B26877.
 DR PIR; A26877; A26877.
 DR PIR; S48478; S48478.
 DR SCD; S0001458; MUC1.
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KW Signal; Multigene family. POTENTIAL.
 FT SIGNAL 1 21
 FT CHAIN 22 1367 GLUCOAMYLASE S1/S2.
 FT DOMAIN 210 1367 SER/THR-RICH.
 FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SEQUENCE 1367 AA; 1361110 MW; 91C00E2DBD61AA9D CRC64;
 Query Match 12.0%; Score 789.5; DB 1; Length 1367;
 Best Local Similarity 28.4%; Pred. No. 7.9e-26;
 Matches 301; Conservative 104; Mismatches 483; Indels 171; Gaps 47;
 QY 59 TTPDTSTTQHNVKSTSPKITTAKPINP-----RPSLPNSDTSKETSILVNKETTVE 110
 Db 236 STSESTTTSTSTSESTSSSTTAPATPTTCTCKEKPPTTTSCTCKEKPPTPHHDTPC 295
 QY 111 TKETTITNKQSTDCGKEKTSKAKQTSKADLAPTSKVLAKPPKAEKTT--TKGPA 168
 Db 296 TKKTKTSK-TCT---KKTTPVPVTPS-SSTESSAEV-----PTSSSTTESSAPV 344
 QY 169 LTPKE-----PTPTTKEPASTTPKEPTPTTIKSAP---TTPKEPAPTTTTSAPTTP 218
 Db 345 TSSTESSAPVPTPSSSTSSAPVTSSTESSAPVTSSTESSAPVPTPSSSTTE 404
 QY 219 KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 278
 Db 405 SSSAPVTS---TTESSAPVTS---TTESSAPVTS---TTESSAPVTSSTTES 453
 QY 279 TTPKEPTP---TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 331
 Db 454 SSAPVPTPSSSTTESSAPVPT---SSTESSAP-VPTPSSSTTESSAPVTSSTESS 509
 QY 332 APITTKPEPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 387
 Db 510 APVPTPS-SSTESSAPAPVPTPSSSTTESSAPVTSSTESSAPVPTPSSSTTES 568
 QY 388 ---TTPKEPAP-TPKKAPVPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 442
 Db 569 VTSSTESSAPVPTPSSSTTESSAPVPT---PSSSTESSAPAPVPTPSSSTESSA 624
 QY 443 PTPPKLTPPTPEKLPATTPKEPAPTTPEELAPT-TPEEPTTTPPEEPAP-----TTPKA 496
 Db 625 PVT---SSTESSAP-VPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTES 679
 QY 497 AAPNTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 555
 Db 680 SAPVT---SSTESSAPVPT---SSTESSAPVPTPSSSTTESSAPVPTPSSSTTE 731
 QY 556 KELAPTTTKEPTSTSDKAPVPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 611
 Db 732 SSGAPVPTPS-SSTESSAPVPTSSSTESSAPVPTPSSSTTESSAPVPTPSSSTTES 790
 QY 612 KEPAPTTTKEPAPVPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 666
 Db 791 SAPVPTPSSSTTESSAPVPTPSSSTNITSSAPVPTPSSSTTESSAPVPTPSSSTTES 850
 QY 667 PAPVPTTTPPTTSEVSTPT---TTPKEPTTIHKSPDESTPELSAPPTPKALENSPKPEP 721
 Db 851 SAPVSSSTTESSAPVPTPSSSTNITSSAPVPT---PFSSTTESFSTGT-TVTPSSSKYP 906
 QY 722 GVPVTKPAATKPEMTTATKDKTTERDLRTPTTTPAAPKMTKETATTTTEKTESKITAT 781
 Db 907 GSQTEISVSSSTETIIVPTKTTTTPPTTTTTPPTTTTTPPTTTTTPPTTTTTPPTTTTTP 966
 QY 782 --TQVVTSTTTQDTPPKITTLTKTTTLAPKVTT---TKKTIITTT-ELMNKPEETAKPKDR 835


```

Db 967 VPTTTTSTSTTTITTTCTGCTNSAGETTSCSKPTITTTVPCTSTSEFA----- 1021
QY 836 ATNSKATPKPKPK-----TKAPKPTSTKKPKTMVRKPKTTTPR 878
Db 1022 ---SBSTTSPPTTPTVVSTVVVTEYSTSTKPGGEITTFVTKNIPTTTLTTIATTP- 1077
QY 879 KMTSTWPELNPSTRIAEAMLOTTTRPNOTPNSKLVEVNPKESEDAGAGETPHMLLRPHV 938
Db 1078 -SVTIVTFTPT-ITTTVCSTGT-----NSAGETSGGSPKRTVTTVPCTST 1122
QY 939 FMEVTPDMDLPRVPNGIINPMLSDETNICNGKPDVGLTTLRNGTLVAFRGHYFWML 998
Db 1123 GTGEYTTAATLVTTAVTTTAVTTTVESSGTNSA-GKTTTGTTKSVPTT-----YVTTL 1175
QY 999 SPFSPSPARRITEWGPSPIDVFFTRC-----NCEGKT 1033
Db 1176 APSAPVTATN-----AVPTTTITT--TECSAATNAAGET 1207

RESULT 4
MUC1_HUMAN
ID MUC1_HUMAN STANDARD; PRT: 1255 AA.
AC P15941; P15942; P13931; P17636; Q14128; Q16442; Q16437; Q9Y4J2;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)
DE (EPISIALIN) (TUMOR-ASSOCIATED EPITHELIAL MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG) (PEANUT-
DE REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN
DE DF3).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=90368716; PubMed=2394722;
RA Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9020794; PubMed=2318825;
RA Lightberg M.J.L., Vos H.L., Gennissen A.M.C., Hilken J.;
RT "Episialin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini.";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;

```

```

RA Wreschener D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,
RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,
RA "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschener D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transfected gene and over-expression in breast cancer
RL tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]
RP SEQUENCE FROM N.A.
RX MEDLINE=91033045; PubMed=1688329;
RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschener D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RL Gene 93:313-318(1990).
RN [8]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88330762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [9]
RP SEQUENCE OF 1-169 FROM N.A.
RX MEDLINE=90088473; PubMed=2597151;
RA Abe M., Siddiqui J., Kufe D.;
RT "Sequence analysis of the 5' region of the human DF3 breast
RT carcinoma-associated antigen gene.";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [10]
RP SEQUENCE OF 1-109 FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschener D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene.";
RL Int. J. Cancer 66:55-59(1996).
RN [11]
RP SEQUENCE OF 1-89 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.-J., Yang P.-C., Shew J.-Y., Hong T.-M., Yang S.-C., Lee Y.-C.,
RA Lee L.-N., Luh K.-T., Wu C.-W.;
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT tissues.";
RL Oncology 53:118-126(1996).
RN [12]
RP SEQUENCE OF 1-46 FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [13]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;
RA Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [14]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;

```

CC VARIES FROM 21 TO 125 IN THE NORTHERN EUROPEAN POPULATION. THE
CC MOST FREQUENT ALLELES CONTAINS 41 AND 85 REPEATS.
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J05582; AAA60019.1; -
CC EMBL; M32738; AAA35804.1; -
CC EMBL; M32739; AAA35806.1; -
CC EMBL; J05581; AAA59876.1; -
CC EMBL; M61170; AAB53150.1; -
CC EMBL; X52229; CAA36477.1; ALT_SEQ.
CC EMBL; X52228; CAA36477.1; ALT_SEQ.
CC EMBL; M35093; AAB59612.1; ALT_SEQ.
CC EMBL; Z17324; CAA78972.1; -
CC EMBL; Z17325; CAA78973.1; -
CC EMBL; M31823; AAA35757.1; -
CC EMBL; S81781; AAD14376.1; ALT_INIT.
CC EMBL; S81736; AAD14369.1; ALT_INIT.
CC EMBL; M21868; AAA59874.1; ALT_SEQ.
CC PIR; A35175; A35175.
CC PIR; B35175; B35175.
CC PIR; S10218; S10218.
CC GlycoSuiteDB; P15941; -
CC MIN; 158340; -
CC MIN; 113720; -
CC InterPro; IPR000082; SEA.
CC Pfam; PF01390; SEA; 1.
CC SMART; SM00200; SEA; 1.
CC PROSITE; PS00024; SEA; 1.
CC Glycoprotein; signal; Cytoskeleton; Actin-binding; Transmembrane;
CC Repeat; Alternative splicing.
CC SIGNAL 1 23
CC CHAIN 24 1255
CC DOMAIN 24 1162
CC TRANSMEM 1163 1186
CC DOMAIN 1187 1255
CC DOMAIN 81 960
CC DOMAIN 1034 1151
CC CARBOHYD 957 957
CC CARBOHYD 975 975
CC CARBOHYD 1029 1029
CC CARBOHYD 1055 1055
CC CARBOHYD 1133 1133
CC CARBOHYD 19 19
CC VARSPLIC 20 22
CC VARSPLIC 20 31
CC VARSPLIC 126 905
CC VARSPLIC 1077 1087
CC VARSPLIC 1088 1255
CC CONFLICT 2 2
CC CONFLICT 134 134
CC CONFLICT 154 154
CC CONFLICT 1021 1021
CC CONFLICT 1251 1251
CC SEQUENCE 1255 AA; 122072 MW; 5E28DFC4DE7D9A82 CRC64;
CC
CC Query Match 9.9%; Score 651; DB 1; Length 1255;
CC Best Local Similarity 27.8%; Pred. No. 3.3e-20;
CC Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps 51;
CC
CC 99 TSLTVNKEITVETKETTNTKNTSGDKKTTSAKETOSIEKTSARD-LAPTSKYLAKPT 157
CC 16 TVLTV-----VTGSGHASTPGGKEKTSATQSRSSVPSSTKNAVSMTSSVLSSHS 65
CC
CC

QY 158 P-KAEITTKGP--ALTTPKEP-----TPTTPKEPASTTPKPEPTPTI--K 197
DB 66 PGSSSTTQGDVTLATAPETASGSAATWGODVTSVPTPALGSGTTPPAHDVTSAPDNK 125
QY 198 SAPTTTPKEPATTTKSAPTTTPKEPATTTPKEP-----APTTPKEPATTTPKEPATTTK 251
DB 126 PAPGSTAPPAGHVT--SAPDT--RPAGSTAPPAGHVTSAFDT--RPAGSTAPPAGHVT-- 179
QY 252 SAPTTTPKEPATTTP-----KKPAPTTPKEP-----APTTPKEPATTTP----- 289
DB 180 SAPDTRPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVT 239
QY 290 ----KEPATTTPKEP-----APTTPKEPATTTPKEP-----KAPTTPKEPATTTPKEP-----E 330
DB 240 SAPDTRPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVT 299
QY 331 PAPTTTKPEPSTTPKEPATTTP-----TKSAPTTTPKEPATTTKSAPTTTPKEPSTTPKEP-- 385
DB 300 SAPDTRPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVT--SAPDT--RPAGSTAPPAGH 356
QY 386 ----APTTPKEPATTTPKEP-----KAPTTPKEPATTTPKEP-----EPAPTTPKEPATTTPKEP 433
DB 357 GVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGH 416
QY 434 APTTPKEPATTTPKKTTPPTPEKLAAPTTPKEPATTTPPEELAPTTPPEEPTTPPEPAPT 493
DB 417 GVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGH 472
QY 494 PKA-----AAPNTTPKEPATTTPKEP-----EPAPTTPKEPATTTPKEPATTTPKEPATTTPKEP 545
DB 473 PPAHGVTSAPDTRPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGH 528
QY 546 APTTP-----KKPAPKELAPTTPKKEPTSTSDKAP--TTPKGTAPTTPKEPAPT 593
DB 529 GSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGH 583
QY 594 TPKEPATTTPKEP-----TAPTTLKEPATTTP-----KKPAPKELAPTTPKKEPTSTT 639
DB 584 TRPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGH 642
QY 640 SKAPAP--TTPK-----ETAPTTPKEPATTTP-----KKPAPTTPETPPPTTSEVS 683
DB 643 DTRPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGH 700
QY 684 TPPTTKKEPT-----IHKSDESTPELSAETPKALENSKEPCVPTTKTAPAKKPEM 736
DB 701 APDTRPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGH 753
QY 737 TTTAKDKTTERDLRTTPETT-----TAAPKMTKETATTTTEKTESKITATTQTSTT 789
DB 754 --PAHGVTSAPDTRPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGH 807
QY 790 TODTTP--FKITLTKTTTLAPKVTTTKKTIITTT--EIMNKPETAKPKBRANSKATTPK 845
DB 808 PGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGH 867
QY 846 PQKPTKAPKPTTKKPKMTPRVKPKTTPTRKMTSTMPELNTPSKTAEAMLOTTTRPN 905
DB 868 P--GSTAPPAGHVTSAFDTTP-----AFGSTAPPAGHVTSAFDTTPA-----PG 909
QY 906 QT--PNSKLVENPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPFRVNPQIINPML 964
DB 910 STAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGSTAPPAGHVTSAFDTTPAGH 955
QY 965 SDETNCNGRPVDGLTLRLNGTLVAFRGHYFWMLSPSPSPS 1005
DB 956 HNVTSAGSAGSASTLVLHNGTSARATTTTPASKSTPESIPS 996

RESULT 5
FPL_MYTED
ID FPL_MYTED STANDARD; PRT; 875 AA.

AC Q25460;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE ADHESIVE PLAQUE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT
 DE PROTEIN 1) (MEFPI) (FRAGMENT).
 GN FPI.
 OS Mytilus edulis (Blue mussel).
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
 OC Mytilidae; Mytilidae; Mytilus.
 OX NCBI_TaxID=6550;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91025829; PubMed=1367451;
 RA Filipula D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;
 RT "Structural and functional repetition in a marine mussel adhesive
 protein";
 RL Biotechnol. Prog. 6:171-177(1990).
 RN [2]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
 RX MEDLINE=83135732; PubMed=6298211;
 RA Waite J.H.;
 RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
 hydroxyproline-containing, decapeptide in the adhesive protein of the
 mussel, Mytilus edulis L.";
 RL J. Biol. Chem. 258:2911-2915(1983).
 CC -|- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
 PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S
 ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
 FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
 CC -|- SUBCELLULAR LOCATION: SECRETED.
 CC -|- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
 CC -|- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
 CC -|- PWM: THE DECAPEPTIDE A-K-P-S-Y-P-T-Y-K IS POST-TRANSLATIONALLY
 MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
 HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
 (DOPA) DERIVED FROM TYROSINE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54422; CAA38294.1;
 DR InterPro: IPR002964; Adhesive_plaq.
 DR PRINTS: PR01216; ADHESIVEI.
 DR PRINTS: PR01217; PRICHEXTENSN.
 KW Repeat; Hydroxylation.
 FT NON_TER
 FT DOMAIN 67 870 TANDEM REPEATS OF Y-K-[PA]-K-[LP]-[ST]-Y-
 P-P-[ST].
 SQ SEQUENCE 875 AA; 100412 MW; 6EA85312748CAACE CRC64;
 Query Match 8.5%; Score 556.5; DB 1; Length 875;
 Best Local Similarity 28.7%; Pred. No. 1.7e-16;
 Matches 272; Conservative 106; Mismatches 375; Indels 195; Gaps 56;
 QY 66 TOHNKYSTSPKITTAKPINP-----RP--SLPPNSDTSKETSILVTKETTVE 111
 Db 1 TKHEPVKPKVSYSAPYKPPYQPLKKVDYRPTKSPPTVG-SKTNVLPKAKLSYK 59
 QY 112 KETTINKOTSDG--KEKTT---SAKETQSIKTSKADLAPTSKVLAKPTPKAETTKG 166
 Db 60 PIKTTYNAKNTPPVYKPKMTPPTTYKPKSPPTTYKSKPTVKKITIYPPYKAKPSY-- 117
 QY 167 PALTTPKEPTTPKE---PASTTPKEPTPTTIKSAETTP-----KEPAPTTKSAPTT 217
 Db 118 PPSYKPKTYPTYPKPLTYPTYPKPKSPPTTYKPKSPPTTYKPKSPPTTYKPKSP 177

QY 218 PKEPAPTTTKEPA--PTTPKEPA-----PTTTKEPA--PTTTKSAPTTPK--EPAPTT 264
 Db 178 P-----PTTKAKPSYPTTYKAKPSYPTTYKAKPTTYPSYKAKPSYPTTYKAKPTTY 233
 QY 265 PKPA--PTTPKEPA--PTTPKEPT--PTTPKEP-----APTTPKEPAPTTPKEPA--PTAPK 313
 Db 234 KAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPTTYKAKPTTYKAKPTTYKAKPSYPTTYKA 293
 QY 314 KPA--PTTPKEPA--PTTTKEPS--PTTTKEPS--PTTPKEPA--PTTTKSAPTTPKEPA-- 363
 Db 294 KPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPTTYKAKPTTY 353
 QY 364 PTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTPKPA--PTTPKEPA--PTTPKEPAPT 419
 Db 354 PSTYKAKPSYPTTYKAKPSYPTTYKAKPTTYKAKPSYPTTYKAKPSYPTTYKAKPSY 409
 QY 420 TTTTKPAPTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPKEPAPTTPBELAPTTP 479
 Db 410 PPTTKAKPTTYKAKP-----TYPST-YKAKPSYPTTYKAKPSYPTTYKAKPTTYKAKPTTYKA 457
 QY 480 EPTPTTPPEPAPTTPKAAAPNTPKPAPTTPKEPAPTTPKE-----PAPTTPKETAPTTP-- 533
 Db 458 KPT-----YFSTYKA-----KPSYPASAKPSYPTTYKSKSSYPSYKPKKTIPTTYK 506
 QY 534 PKGTAPTTLKEPAPTTPKAPKELAPTTPKPTSTTSKAPAPTTPKGTAPTTP-----PKE 589
 Db 507 PKLTYKPTTYK-BKSPSPSYKPKTTPPTTYK-----PKISYPTTYKAKPSY 551
 QY 590 PAPTTPKEPAPTTPKGT--APTTLKEPA--PTTPKPA--PKELAPTTPKGT--PTSSTSD 641
 Db 552 PATYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTY 611
 QY 642 KPAPTTPKETAPTTPKEPA--PTTPKPA--PT-----PETPPTTSEVSTTTTKE----- 690
 Db 612 KAKPSYPTTYKAKPSYPTTYKAKPTTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTYKAKPSY 667
 QY 691 PTTIHKSDESPELASPTPKALENSPK---BEGVPTT--KTPA-----ATKPEMTTT 739
 Db 668 PPT-YKAKPSYPTTYKAKPTTYKAKPTNPSTYKAKPSYPTTYKAKPSYPTTYKAKPSYPTTY 726
 QY 740 AKDKTTERDLRTPPTTTTAAKPKMTKETATTTKTESKITATTTQVSTSTTTQDTPPKIT 799
 Db 727 YKAKPTTYKAKPTTYKAKP--TYKAKPTTYPTTYKAK-----PSYPTTYKPKPSYPTTY 777
 QY 800 TLKTTTLAPKVTITTKITTTTTEIMNKPETAKPKDRATNSKATTPKPKQPKAPKK---P 856
 Db 778 TYKSKSIYPSYKPKKTYPTTP---YKPKLTYPTTYK-----PKPSYPTTYKPKITYP 826
 QY 857 TSTKPKTKMPRVKPKTTPTP-----RKMTSTMPELNPTSR 892
 Db 827 STYKLPKSPPTTYKSKTSYPTTYKSKTSYPTTYKSKTSYPTTYKSKTSYPTTYKSKTSYPTTY 874

RESULT 6
 ZAN_HUMAN

ID ZAN_HUMAN STANDARD; PRT; 2700 AA.

AC Q9Y493; Q00218;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE ZONADHESIN (FRAGMENT).

GN ZAN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE OF 1-2379 FROM N.A.

RX MEDLINE=99018118; PubMed=9799793;

RA Glockner G., Scherer S., Schattevoy R., Boright A., Weber J.,

RA Tsui L.C., Rosenthal A.;

RT "Large-scale sequencing of two regions in human chromosome 7q22:

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV; XANTHI; TISSUE=Leaf;
 RX MEDLINE=90128263; PubMed=2612909;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich
 glycoprotein gene in lateral root initiation.";
 RL Genes Dev. 3:1639-1646(1989).
 CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
 THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
 MAIN ROOT.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
 CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 GLYCOSYLATED.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X13885; CAA32090.1; -;
 DR PIR; S06733; S06733.
 KW Hydroxylation; Cell wall; Glycoprotein; Signal; Structural protein;
 FT SIGNAL
 FT CHAIN 1 ?
 FT REPEAT 7 620 EXTENSIN.
 FT REPEAT 70 73 H-A-P-P.
 FT REPEAT 148 151 H-A-P-P.
 FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 229 235 1.
 FT REPEAT 236 242 2.
 FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
 FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
 FT SEQUENCE 620 AA; 63406 MW; 641DD2278AB28524 CRC64;

Query Match 8.4%; Score 551; DB 1; Length 620;
 Best Local Similarity 27.2%; Pred.No. 2.1e-16;
 Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;
 QY 160 AETTK-GPALTTP--KEPTTT-----PKPASTTPKE---PTPTTKSAPTTKPEAP 208
 DB 24 AEATQYGGYLLPPVTSQPPSSIGLSPSPSAPTTTPSRGHVSP---RHAPRHAYPPP 80
 QY 209 TTTKSAPTTKPEP-----APTTPKEPAT---TPKEPATTTKPEAPTTKSAPTTKPE 259
 DB 81 SHGHLPSVGGPPRHGLPDSGFPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 140
 QY 260 PAPTTPKKPAPT---TPKEPATTTKPEAPTTKPEAPTTK---EPATTPKPEAPATPKK 314
 DB 141 P-----PSPSHGAPPSSGGHTPPRGQHPSPSHRSPSRHCHGHPPTTTPYAPPTT 193
 QY 315 PAPTTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTK 369
 DB 194 PSQV--QPPT--YSPPTTHVQTPSPSPSRGHPQPPHAPTHRHAPTHHAPTHHAP 249
 QY 370 ---APTTPKPSPTTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 423
 DB 250 RHLPPSPRQOPPTTSPPPAYAQSPQSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 309
 QY 424 KPAPT---APKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 473
 DB 310 PPTPTFTFSPPPAYSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 369
 QY 474 APTTPEPTPTTTPPEAPTTPKAAP---NTPKEPATTTKPEAPTTKPEAPTTKPEAPTT 530

Db 370 PPPPPSPPPSP 427
 QY 531 PTTTPKGTAPTTLKEPATTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 590
 Db 428 P-----LPPTYSPPPPAYSPPPPTTSPPPPTTSPPPPAYAQPPPPPTTSPPP 481
 QY 591 APTTPKEPATTTKGTAPTTLKEPATTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 650
 Db 482 PPSPIYSPPPQVQLPPTTSPPPRRHLPPPPHQP---RPPTTGYQPPSPPTTSP 538
 QY 651 TAPTTPKEPAT---TPKPEPATTTTTPPTTSEVSTPTTKEPTTIHKSPDESTPELSAE 708
 Db 539 PPRQIHSPPPPHQWQRTTPTTYGQPPSPPTFSAPPRQIHSPPPHQRPPPTTPTTYG 598
 QY 709 PTPKALENSPKPGVPTTKTPAATKP 734
 Db 599 PSP-----PTTYSPPSPPP 612

RESULT 8
 NFH_MOUSE
 ID NFH_MOUSE STANDARD; PRT; 1087 AA.
 AC P19246; O61959;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
 GN NFH OR NFH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89121513; PubMed=3220257;
 RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosfeld F.,
 RA Mushynski W.;
 RT "Sequence and structure of the mouse gene coding for the largest
 neurofilament subunit.";
 RL Gene 68:307-314(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89089138; PubMed=3145094;
 RA Shneiderman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
 RT "The structure of the largest murine neurofilament protein (NF-H) as
 revealed by cDNA and genomic sequences.";
 RL Brain Res. 464:217-231(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
 RA Carden M.J.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPETIDE K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
 CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M24496; AAA39813.1; JOINED.
EMBL; M23349; AAA39813.1; JOINED.
EMBL; M24494; AAA39813.1; JOINED.
EMBL; M24495; AAA39813.1; JOINED.
EMBL; M35131; AAA39809.1; AUT_FRAME.
EMBL; M31012; CAA83229.1; -
PIR; J03368; QPM5H.
PIR; A43778; A43778.
MGD; MGI:97309; NFH.
InterPro; IPR001664; IF.
Pfam; PF00038; filament; 1.
PROSITE; PS00226; IF; 1.
Intermediate filament; Coiled coil; Neurone; Phosphorylation;
Repeat.
DOMAIN 1 97 HEAD.
DOMAIN 98 408 ROD.
DOMAIN 409 1087 TAIL.
GLU-RICH (ACIDIC).
50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.
GLO/LYS-RICH.
COIL 1A.
LINKER 1.
COIL 1B.
LINKER 12.
COIL 2A.
LINKER 2.
COIL 2B.
K -> QA (IN REF. 2 AND 3).
A -> AR (IN REF. 2 AND 3).
S -> T (IN REF. 2 AND 3).
L -> G (IN REF. 2 AND 3).
P -> PREAKSP (IN REF. 3).
MISSING (IN REF. 3).
G -> A (IN REF. 3).
V -> M (IN REF. 2 AND 3).
T -> N (IN REF. 2 AND 3).
SEQUENCE 1087 AA; 116612 MW; 57BAC76A38EDICB9 CRC64;

Query Match 8.1%; Score 533; DB 1; Length 1087;
Best Local Similarity 27.5%; Pred. No. 1.9e-15;
Matches 216; Conservative 81; Mismatches 330; Indels 158; Gaps 38;
QY 45 DEAGSLDNGDFKVT-----TPDTSTQHNKYSTSPKITTAKPINRPSLPNSDTSKET 99
DB 404 EECRIGFSPSLTEGLPKIFISIT--HIKVKSEMIKVV-----KSEKET 449
QY 100 SLTVNKETVETKTTTNNKQSTGKETTSAKETQSIKTSKDLAPTSKVLAKPTPK 159
DB 450 VIVEGQTEIRVETGVEEDKEAEOGEEAEGBEEKEELEAAATSPPAEAAASPK 509
QY 160 AETTTKGPALTTPKEPTTPKEPASTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTPK 219
DB 510 TKSRRVEEAKSPGEAKSPGEAKSPA-----EAKSPGEAKS-PGEAKSPGEAKSPAEPKSPA 564
QY 220 EP-APTTPKEPAPTTKEPAPTTKEP-----APTTPKSAPTTPKEP-----PTTPKKA- 269
DB 565 EPKSPAEAKSPA--EPKSPA--TVKSPGEAKSPSEAKS--PARAKSPAEPKSPAEPKSPA 619
QY 270 ----PTTPKEP-----PTTPKEP-----TPTTPKEPAPTTKEPAPTTKEP-----PTAPK 314
DB 620 AKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 676
QY 315 P-----APTTPKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTTKEPAPTT 366
DB 677 PGEAKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 729

QY 367 TKSAPTTPKE-PSPTTTKEPA-----PTTPKEPAPTTPKKPAPTTPKEPAPTTTPKEPAPTT 421
DB 730 SPAVAKSPGSAKSPGSAKSPGSAKSPGSAKSPGSAKSPGSAKSPGSAKSPGSAKSPGSAKSP 789
QY 422 TTKPAPTAPEPAPTTTPKEPAPTTTPKKTTPKLTTPTEKLAPTTPKEPAPTTPEELAPTTPEEP 481
DB 790 AKSPVKEDIKPPAEAKSPEKA--KSPVKEGAKPPEKAKPLDVKSPAEQTPVQEEATVPTDI 848
QY 482 TPTTPEE-PAPTTPKAAAPNTPKPEPAPTT-----PKPEPAPTT-----KEPAPTTPKET 529
DB 849 RP--PEQVKSAPKAKSPE--KEEATSKVAPKKEEVKSPVKEEVKAKPEPKVKEEEK 904
QY 530 APTTPKGTAPTTLKEPAPTTTPKPAKELAPTTTKEPTSTSD--KPAPTTPKGTAPTT 587
DB 905 TLTPTKEAKESKDEAPKAPKPEKKEPTPEKPKDSTAEAKKEAGEKKAVASEE 964
QY 588 KEAPTTTPKEPAPTTPKGTAPTTLKEPAPTTPKKPAKELAPTTTKEPTSTSD--KPAPTTPKGTAPTT 647
DB 965 ETPAKLGVKEEA--KPEKETTETTKTEADTKAKPS-----KPTETEP-----1006
QY 648 PKETAPTTPKPEPAPTTPKKPAPTTPPETPPPTTSEVSTPTTKEPTTHKSPDESTPELSA 707
DB 1007 -----KKEEMPAPEK-----KDTKEEKTIESRKPEEK-----1034
QY 708 EPTPKALENSPKPGVPTTKTAPKPKEMTTAKDKITERDLRTTPTTTAAKPKMTKETA 767
DB 1035 ---PR-MEAKVKEDDKSLSKEP--SKPKTEKAESSSTDKESOPPE-----KTTEDKA 1082
QY 768 TTTTEK 772
DB 1083 TKGEK 1087
RESULT 9
TCNA_TYCR STANDARD; PRT: 1162 AA.
ID TCNA_TYCR
AC P23253, 1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE STALIDASE (EC 3.2.1.18) (NEURAMINIDASE) (NA) (MAJOR SURFACE ANTIGEN).
GN TCNA.
OS Trypanosoma cruzi.
OC Eukaryota; Euclenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID:5693;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SILVIO X-10/4;
RX MEDLINE-91277609; PubMed-1711561;
RA Pereira M.E.A., Mejia J.S., Ortega-Barria E., Matzilevich D., Prioli R.P.;
RA "The Trypanosoma cruzi neuraminidase contains sequences similar to bacterial neuraminidases, YWTD repeats of the low density lipoprotein receptor, and type III modules of fibronectin.";
RL J. Exp. Med. 174:179-191(1991).
RN [2]
RP SUBCELLULAR LOCATION.
RX MEDLINE-91376547; PubMed-1896773;
RA Prioli R.P., Mejia J.S., Aji T., Alkawa M., Pereira M.E.A.;
RT "Trypanosoma cruzi: localization of neuraminidase on the surface of trypomastigotes.";
RL Trop. Med. Parasitol. 42:146-150(1991).
CC -1- FUNCTION: DEVELOPMENTALLY REGULATED NEURAMINIDASE IMPLICATED IN PARASITE INVASION OF CELLS.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 2,3-, 2,6- AND 2,8-GLYCOSIDIC LINKAGES JOINING TERMINAL N-REDUCING N- OR O-ACYLEURAMINYL RESIDUES TO GALACTOSE, N-ACETYLHEXOSAMINE, OR N- OR O-ACYLEURAMINYL RESIDUES IN OLIGOSACCHARIDES, GLYCOPROTEINS, GLYCOPOLIPIDS OR COLOMINIC ACID.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POSSIBLE).
CC -1- DEVELOPMENTAL STAGE: MAXIMAL ACTIVITY IN TRYPOMASTIGOTES, MINIMUM IN EPIMASTIGOTES AND NO DETECTABLE IN AMASTIGOTES.


```

Qy 11 LLLSVFVLIQQVSVKDNKNKRTKPKPPVVDGAGLDNGDFKVTTPTDST--TQH 68
Db 8 LCLLCLFSCDIFALSGNIHN-----VYGSAYSASGAYK-TLPGSHPYGSKH 55
Qy 69 -----NKVSTSPKITTAKPINPPSLPPNSDTSKETSITLVNKETTVETKETITNKOT 121
Db 56 VPVYKPMNKIPT-PYI--SKRSYPAPYKPGYYPTKRYQPTYGSKNYPPIYKPIAKKUS 112
Qy 122 STDGKEKTTISAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPT-PTTP 180
Db 113 SYKAIKTTVPYAKAKTSYDPSVK-----HKITYPTVK-----PKITYP--PTYKQP 158
Qy 181 KEPASTTPKEPTPTTIKSAPTTPKBPATTTKSAPT-TPKEPAPT--TTKEPAPTTPEP 237
Db 159 SYPSYKPKITYPTVK----PKITYPTVKRPSVTPYKPKATYPTTKKITYP---- 210
Qy 238 APTTTKEPAPTTTKSAPTTPKBPATTT--PKKAPTTTPKEPAPT--PKEPTPTT--P 289

```

Db 159 SYPSPYKPTTDPYK-----PKTYTPYKRPSTPYKPKATYPPTYKPYTYP--- 210

QY 238 APPTTKPAPPTTKSATPTTKPAPPT--PKKPAPPTPKPAPPT-----PKETPTPT--P 289

567 LPQVADCKVAFSSR 580

RESULT 13

RPBL_HUMAN	STANDARD;	PRT;	1970 AA.
P24928;			
01-MAR-1992 (Rel. 21, Created)			
01-MAR-1992 (Rel. 21, Last sequence update)			
20-AUG-2001 (Rel. 40, Last annotation update)			
DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RPBL).			
POLR2A.			
Homo sapiens (Human)			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
NCBI_TaxID=9606;			
[1]			
SEQUENCE FROM N.A.			
MEDLINE=92178992; PubMed=1542581;			
Winczler M., Aker J., Vicaire S., Vigneron M., Keding C.;			
"Complete sequence of the human RNA polymerase II largest subunit.";			
Nucleic Acids Res. 20:910-910(1992).			
[2]			
SEQUENCE FROM N.A.			
MEDLINE=95347616; PubMed=7622068;			
Mita K., Tsuji H., Morimyo M., Takahashi E., Nenoi M.,			
Ichimura S., Tamauchi M., Hongo E., Hayashi A.;			
"The human gene encoding the largest subunit of RNA polymerase II.";			
Gene 153:285-286(1995).			
-!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION			
OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS			
SUBSTRATES.			
-!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +			
RNA(N).			
-!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.			
-!- SUBCELLULAR LOCATION: NUCLEAR.			
-!- PTM: THE TANDEM 7 RESIDUE REPEATS CAN BE HIGHLY PHOSPHORYLATED.			
-!- THE PHOSPHORYLATION ACTIVATES POL2.			
-!- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE			
FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA			
PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE			
III FOR 5S AND TRNA GENES.			
-!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.			
THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
between the Swiss Institute of Bioinformatics and the EMBL outstation			
the European Bioinformatics Institute. There are no restrictions on its			
use by non-profit institutions as long as its content is in no way			
modified and this statement is not removed. Usage by and for commercial			
entities requires a license agreement (see http://www.isb-sib.ch/announcement/			
or send an email to license@isb-sib.ch).			
EMBL; X63564; CAA45125.1; -			
EMBL; X74874; CAA52862.1; -			
EMBL; X74873; CAA52862.1; JOINED.			
EMBL; X74872; CAA52862.1; JOINED.			
EMBL; X74871; CAA52862.1; JOINED.			
EMBL; X74870; CAA52862.1; JOINED.			
PIR; S21054; S21054.			
MIM; 180660; -			
InterPro; IPR000684; RNA_POL_II_repeat.			
InterPro; IPR007722; RNA_POL_A.			
InterPro; IPR002879; RNA_POL_A2.			
Pfam; PF00623; RNA_POL_A; 1.			
Pfam; PF01854; RNA_POL_A2; 1.			
PROSITE; PS00115; RNA_POL_II_REPEAT; 43.			
Transferrase; DNA-directed RNA polymerase; transcription; zinc; Repeat;			
DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.			
C2H2-TYPE (POTENTIAL).			
ZN-FING 71 87			
DOMAIN 1590 1958			
CONFLICT 1067 1067			
CONFLICT 1449 1449			
W -> L (IN REF. 2).			
D -> Y (IN REF. 2).			
CARBOXYL-TERMINAL 7-RESIDUE REPEATS.			


```
Db 412 KSPVEAKSPAENVKSPVTVKSPAENVKSPA-SVKSPSEAKSPAGAKSPAENKSPVV 470
QY 312 PKKPAPTTPKKPAPTTPKKPAPTTTKESPPTPKPEA-----PTTKS-----APTTHKEP 362
Db 471 AKSPA--KSPAGAKSPAENKSPAENKSPAENKSPAENKSPAENKSPAENKSPAENKSPA 528
QY 363 APTTKS-----APTTPKEPSPTTTKEPA-----PTTPKEPAPTTPKKPAPTTPKPEAPT 413
Db 529 VKGAKSLAEAKSPEKAKSPVKEEKPPAEVKSPEKAKSPMKKEAKSPEKAKTLDVKSPE 588
QY 414 PKEPAPTTTKKPAPTKAPKEPAPTTPKETAPTTPKLTTTPEKLPAPTTPEKAPAPTTPEEL 473
Db 589 AKPPAKEEAKRPADIRSPQVKSPEKEAKSPEKEETRT--EKVAPKKEEVKSPV--EEV 644
QY 474 APTTPEPTTTPEEPAPTTPKAAAPNTPKPEAPTTPKPEAPTTPKPEAPTTPKETAPTT 533
Db 645 ---KAKEPPKYEEETPATPKTEVKESKKDEAPKEAQP-KAEKEPLTEKPKDSPGEA 700
QY 534 PKGTAPTTLKPEAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKPEAPT 593
Db 701 KKEEA---KEKAAAPBEETPAKLGKVEAKPKKEADAKA-----KEFSKP 744
QY 594 TPKEPAPTTPKGTAPTTLKPEAPTTPKKPAKELAPTTTKEPTSTTSKDPAPTTPKETAP 653
Db 745 SEKE---KPK-----KEVPAPEKKTKE-----EKTESKKREKPKMEAK 784
QY 654 TTPKEPAPTTPKKPAPTTTPETPPPTTSEVSTPTTKEPTTIHKSPDE 700
Db 785 AKED-----KGLPQEPSPKPKTEKAEKSSSTQDKDSQFSEKAPED 824
```

Search completed: April 26, 2002, 16:32:03
Job time: 553 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:17:40 ; Search time 61.21 Seconds
(without alignments)
451.831 Million cell updates/sec

Title: AAL

Perfect score: 6568

Sequence: 1 MAWKTLPIVLLLLLSVEFIQ.....ARAITRSQTLSKVNYP 1229

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	522	7.9	805	4	US-09-103-429A-4
2	508	7.7	1837	3	US-08-928-361B-5
3	506.5	7.7	744	6	5202236-25
4	498.5	7.6	786	4	US-09-103-429A-3
5	489	7.4	1721	3	US-08-700-651-5
6	489	7.4	1721	3	US-08-928-361B-6
7	488.5	7.4	826	1	US-07-538-431-2
8	488.5	7.4	826	5	PCT-US92-00018-2
9	476.5	7.3	652	6	5202236-13
10	452	6.9	960	4	US-09-219-849-5
11	424.5	6.5	1867	2	US-08-479-537A-5
12	424.5	6.5	1867	4	US-09-083-116-5
13	424.5	6.5	2035	2	US-08-479-537A-2
14	424.5	6.5	2035	4	US-09-083-116-2
15	419.5	6.4	2476	2	US-08-276-967-2
16	419	6.4	1185	4	US-09-041-886-23
17	417	6.3	829	1	US-08-642-255-132
18	417	6.3	829	1	US-08-397-633A-53
19	417	6.3	837	1	US-08-175-155-68
20	417	6.3	837	1	US-08-477-509B-103
21	417	6.3	837	1	US-08-642-255-101
22	417	6.3	837	2	US-08-707-237A-75
23	417	6.3	837	3	US-08-482-085B-103
24	417	6.3	897	1	US-08-397-633A-50
25	413.5	6.3	907	3	US-08-789-774-2
26	413.5	6.3	907	5	PCT-US95-04611A-19
27	404.5	6.2	408	1	US-07-609-716-65

28	404.5	6.2	408	4	US-08-475-411A-65
29	404.5	6.2	408	4	US-08-478-029A-65
30	398.5	6.1	682	1	US-08-642-255-126
31	398.5	6.1	682	1	US-08-397-633A-36
32	381	5.8	1848	4	US-08-296-791-6
33	381	5.8	1848	5	PCT-US95-10661A-6
34	376	5.7	960	4	US-09-219-849-6
35	374.5	5.7	1537	1	US-08-325-263A-2
36	373.5	5.7	1231	3	US-08-904-267A-4
37	371	5.6	761	2	US-08-707-237A-84
38	371	5.6	762	1	US-08-642-255-114
39	371	5.6	762	1	US-08-397-633A-26
40	370.5	5.6	1064	1	US-08-642-255-62
41	368	5.6	762	1	US-08-642-255-120
42	368	5.6	762	1	US-08-397-633A-31
43	367.5	5.6	1187	1	US-08-320-559-28
44	367.5	5.6	1187	3	US-08-545-860D-28
45	367.5	5.6	1187	5	PCT-US94-04436-28

ALIGNMENTS

RESULT 1

US-09-103-429A-4

; Sequence 4, Application US/09103429A

; Patent No. 6187538

; GENERAL INFORMATION:

; APPLICANT: Granados, Robert R

; APPLICANT: Wang, Ping

; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin

; TITLE OF INVENTION: cdna and Related Products and Methods

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.

; STREET: 118 No. 6187558th Tloga

; CITY: Ithaca

; STATE: NY

; COUNTRY: USA

; ZIP: 14850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA: US/09/103.429A

; FILING DATE: 24-JUN-1998

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Michaels, Christopher A

; REGISTRATION NUMBER: 34,390

; REFERENCE/DOCKET NUMBER: BTI-39

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (607) 256-2000

; TELEFAX: (607) 256-3628

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 805 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Trichoplusia ni

; TISSUE TYPE: peritrophic membrane

; US-09-103-429A-4

Query Match

Best Local Similarity 7.9%; Score 522; DB 4; Length 805;

US-09-103-429A-4

Mon Apr 29 08:35:13 2002

Matches 226; Conservative 33; Mismatches 211; Indels 342; Gaps 40;

QY 149 TSKVLAKPTPKAETTTGGALTTTKEPTTTTKEPASTTTPKEPTTTIKSAP-----TTP 203
Db 9 TAILGLVAARPEVSDAEKNALHEPHDPX--PAEQXLLLPXEYDCTKFYCEYGLKFIAP 66
QY 204 KEPAPTT-----TKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 251
Db 67 RDCAPGTFFKFSQAQTCVHAALAGCTLPGPAETT---QAPATTQ---APTITQ---APTITTT 119
QY 252 SATTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTA 311
Db 120 QAPTTTQATTTT-----QAPTTTQATTTT-----QATTTQATTTT----- 156
QY 312 PKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 371
Db 157 ---QAPTTTQ---APTITQ---APTITQ---APTITQ---APTITQ---APTITQAA 198
QY 372 TTPKEPSTTTKEPAPTTT---KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 415
Db 199 TTPAATTAAAT--TPAATTAAATTAATTGVPAPTS-----APVPPICELLNGCPADFD 253
QY 416 -----PA-----EPAPTTT 423
Db 254 IHLPHDKYCNLFYQCSNGYTFEORCEPGLFYFNYYVQRCDSPANVECDGEISPAFPVTE 313
QY 424 K-----PA-----TPAP--KEPAPT-----TPKET 441
Db 314 GNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYQCVHGNLVERRCGAGTHFSFELQ 373
QY 427 -----TPAP--KEPAPT-----TPKET 441
Db 374 QCDHIELVGCPLPGSESEVVDDEACTGWCTEPIEWEPLNGCPADFSIDHLLPHES 433
QY 442 -----APTTPKLTPTT-----PEKLAPTTPEKAP 467
Db 434 DCGQYLVQCVHGOTIARPCGNLHFSPTQSCSPVTAGCVFECDSNOCSTAAAPTAA 493
QY 468 TPEELAPTTTPEETPTTPEPAPTTPKAAAPNTPEPAPTTPKEPAPTTTKEPAPTTT 527
Db 494 TAAPTAAPTAAAPTAASTVWPVA--TPATAAPVPTTAIPT-----PAPTAAAPTAA 548
QY 528 ETAPTT---PKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTSDKAPTTPKGTAP 584
Db 549 PESPTTVVPTTAAPTA---APT---AVPEIPTVTSAPTAAPT---AAPTAAAPTAA 598
QY 585 TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 643
Db 599 TTAAPPEIPTT-----VTSPTTAAPT---AAPTNT-----TVVPTTAAPTAA 641
QY 644 APTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 703
Db 642 AP---NTVTAPPTAAPTT---AAPTNTVTVPTTAAPTAA---PTVAH-----AP 685
QY 704 ELSAAPTTPALENSPKFCVPTTKTTPAATKPE 735
Db 686 NTAAP-----VTTTSAPATT-PE 703

RESULT 2
US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSES: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vermy, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480,76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1837 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-5

Query Match 7.7%; Score 508; DB 3; Length 1837;
Best Local Similarity 21.3%; Pred. No. 4.5e-26;
Matches 295; Conservative 100; Mismatches 494; Indels 496; Gaps 48;

QY 57 KVTPTDSTTQHNKVTSPKITTAKINRPSLPNDSDTSKETSLVKNKETTVEKRETTT 116
Db 307 KHTT 366
QY 117 TNKQTSIDGKEKTSAKETQSIKTSKDLAPTSLVAKRPTKAEETTKGALTTPKEPT 176
Db 367 TTTTTTTTKKPTTKKPT 426
QY 177 PTPKEPASTTKEPPTTIKSAPTTPKEP--APTITKSAPTTPKEPAPTTPKEPAPTTP 234
Db 427 TT 486
QY 235 KEPAPTTTKEPAPTTTKSAPTTPKEPAPTTPKAPPTTPKEPAPTTPKE----- 283
Db 487 TTTATTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTSETSVIKPDEWCWLEKGECE 546
QY 284 -----PTPTT-----PK 290
Db 547 AKGATVYGVIGKDGRIENGMAFTMPNDTHVRFRKVKDVGNTISVRCGKGAGKLEFPD 606
QY 291 EPAPTTKEP-----APTTPKEPA--PTAPKPE----- 315
Db 607 RSLDFTTPPVAGHNSCSIIIVGSGGGKIHVSGYSGKDVSLISAPIQCELFNEVYCDTCT 666
QY 316 -----APTTPKEPAPT-----PKEPAPTTPKEPS--PTTPKEPAPTTPKS- 354
Db 667 AKYCAIHSGVQTSADFVTTTAKETTTTTTGAPGQPTTTTIGSPSKPTTTTTRATTTT 726
QY 355 -----APTTPKEPAPTTPTKSAPTTPKEPSPTTTPKEPAPTTPKEPAPTTPKEPA 410
Db 727 LNPIITTTTQKPTTTTTPKVPKPIATTTTTTLKPIVTTTTTTRATTTTTPVPTT----- 781
QY 411 PTTTKEPAPTTPKAP-----TAPKEPAP----- 435
Db 782 TTTTKRDEMTTTTPLPDIGIEITPIPIEKMLDKYRMIYDYNSSLNLDSDNDEPIPGSQ 841

QY 436 -----TTPKETA-PTTP-----KKL 449
Db 842 AGCIADTSLNLPVQTHKSTGLPDPNVLFPDPKSGNLVHPYNTMTSGLSVSLAAKNL 901
QY 450 T-----PTTEKLAAPTE-----463
Db 902 TVDVTDTYGLPIDTLTGYPDLDPVSLIPFNPETGELDFPISDEIMNGTIAGIVSGISASES 961
QY 464 ----KPAPTPPE-----LAPTT-----PEEPTPTP--EEPAPTPKAAA 498
Db 962 LLSOKSAPIDPATNMVVBEGGLLPATGVMPGLSGSEQPTPFSPEIEDGGIIPPEVAA 1021
QY 499 PNPPEPAPTPKAPAPTPKE-----PA 522
Db 1022 ANADKFKLSIPSPESPEIKDQKIDSISELMYDIESGRLLIGQVSKRPIGPSIAGDLNPI 1081
QY 523 PTPPKETAPTPKGTAPTLKPEAPTPPKKAPKAPKELATTTKEPTSTSDKAPAPTPKGT 582
Db 1082 MKTPQTDSVTKPDKPTT--GLFNP--PTGHLINPTNNMTDSSFAGAYKAVSNGI 1136
QY 583 APTT-----PKAPAPTPKAPAPTPKGTAPTLKPEAPTP-----KKPAPK-- 625
Db 1137 KTDNYVGLPVDSEITGLPKDPVSDIPFNSTTGTGLVDPSTGKPINNYTAGIVSGKGLPIE 1196
QY 626 ----ELAPTTTKGP-----TSTT-----SDKAPAPTPKETAAPTPKPEA--- 660
Db 1197 DENGFLPDSYKLPIDGNNQNLVNETNSTVSGSTGSKPKPGIPVNGGVVDPDEAKDQ 1256
QY 661 -----PTTPKKAPAPTPPTPPPTTSEVSTPTTKE-PTTIHKS---PDESTPEL 705
Db 1257 ADKGRGLIVPTNSINKDPVTNTQYNTIGNIINPETGKVPGLSGSLNYSFNTPOQ 1316
QY 706 SAEPKPALE---NSPKPVGPTTKTAAKPKEMTTAKDTERDLRTPTTETTAAPKM 762
Db 1317 TDEITGKPDVTGLPDPSTGEIIDPATKLPIDPGSVAGDEILTEVLNITTDVETGLP-I 1375
QY 763 TKETATTEKTESKIZATTTQVSTSTTQDTTPKTIITLTKTTLAPKVTTKITTE- 821
Db 1376 DLETGLPRDP-----VSLGPLPGLD-----PSNKKPIFGSHS 1411
QY 822 --INNKPEEAKPDATNSKATPKQKP--TKAP--KKPTSTKKPKTMRVRKPKT--- 873
Db 1412 GFINGTSGEOSHEKDPSTG-----KPLDNTGLPDEDSGLINPETGDKLQSGHSTF 1465
QY 874 TPTPRKMTSTMPENLPTSRIAEAMLQTTTRNPTNSKLVENPK--SEDA----- 922
Db 1466 MPVPGKQGGNGIMPEQILEAL-----NKLPTSNEVNISPRSSDAVDPDRPTNTWW 1518
QY 923 -----GGAGETPHMLL-----RPHVFMPEVTPDMDVLPKRVNPGI 960
Db 1519 NKISGOTFQVDRKKTIPGSAASVIHTALGTPTQDTTGLP--SDPSTGLPIFGENVLV 1576
QY 961 NPMUSD-----ETNIC-----NGKPDVGLTTL---RNGTLV-AF 990
Db 1577 DPQTGEIKGSVPVSVLYKVEKNIVTEAAYGLPVPDKTGFDPIDPSIYLPFAKNGLIDPI 1636
QY 991 RGHYF 995
Db 1637 SGHYF 1641

RESULT 3
5202236-25
PATENT NO. 5202236
APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
SUSAN L.; MCCANDLIS, RUSS; WEI, TENA; FILPULA, DAVID
TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
PROTEIN
NUMBER OF SEQUENCES: 39
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/528,762
FILING DATE: 25-MAY-1990
APPLICATION NUMBER: 82,456

FILING DATE: 07-AUG-1987
APPLICATION NUMBER: 933,945
FILING DATE: 24-NOV-1986
APPLICATION NUMBER: 650,128
FILING DATE: 13-SEP-1984
SEQ ID NO: 25:
LENGTH: 744
5202236-25

Query Match 7.7%; Score 506.5; DB 6; Length 744;
Best Local Similarity 29.9%; Pred. No. 1.9e-26;
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50;

QY 75 PKITTAKPINPRSLPNSDTSKETSILTYNKETTETTTINQOTSTDGKEKTTSAKE 134
Db 24 PKMTYPTTKPKPSYPP-----TKSKPTY-----PKKIT----- 53
QY 135 TQSIETSAKDLAPTSKVLAKTPKPAETTKGPAITTKPEPTTTPKE-----PASTTPKE 190
Db 54 -----YPTVK--AKPS-----YPTVKPKKTYPPPKLTYPPTYKPKP 92
QY 191 PTPPTTKSAPT--TPKEPAPTTTKSAPTTPKEPAPTTTKPEAPTTTPKEAPTTTKEPA--P 247
Db 93 SYPTVKSKPTYPKPLTYPPTYKAKPSYPPTYKPKKTYPP--YKPKLTYPPTYKPKASYP 151
QY 248 TTTKSAPTTPKEAPTPPKKPAPT--TPKEPAPTT-----PKEPTPTTPKEPAPTTKEP--- 299
Db 152 PTYKPKPSYP--PSYKTKATKTPYTPYKPKLTYPPTYKPKPSYPPSYKPKKTYPPTYKPKLT 209
QY 300 -APTTPKEPA--PTAPKPA--PTTPKEPA--PTTPKEPA--PTTPKEPS--PTTPKEPA 348
Db 210 YPTTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPS 269
QY 349 -PTTKSAPTTPKEPAPTTTKSAPTTPKEPSPTTTPKEPA--PTTPKEPA--PTTPKPK 401
Db 270 YPSTYKAKPSYTP---YPSTYKAKPTYP---PTYKAKPSYPPTYKAKPSYPPTYKAKPTYKAKP 322
QY 402 A--PTTPKEPA--PTTPKEPA--PTTPKKA--PTAPKEPAPTTTPKETATTPKLTPTT 453
Db 323 SYPTTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPTYS 382
QY 454 PEKLAPTTPE--KPAPTTPEELAPTTPEEPT--PTTPEEPA--PTTPKA--AAPT---P 502
Db 383 TYKAKPSYPPTYKAKPSYPP---PTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPTYS 438
QY 503 KEAPATTPKEPAPTT-----PKEPAPTTTPKETATTT--PKGTAPTTLKEPAPTTPKKPAK 556
Db 439 SYPTTYKPKISYPTTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKP 498
QY 557 ELAPTTTKEPTSTSDKAPATTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTLKEPA- 615
Db 499 SYPTTYKAKPTYPSTYKAKPTYK--ARPSYPP---PTYKAKPS-----YPPTTYKAKPSY 546
QY 616 -PTTPKPAKELAPTTTKGT--STTSKAPATTPKETATTPKETATTPKEPAPTTTPKKA--PT 670
Db 547 PPTYKAKPTYK--AKPTYKAKPTYPSTYKAKPSYPPTYKAKPSYPP---PTYKAKPSYPP 601
QY 671 ---TPETTPPTTSEVSTPTTKE-----PTTIHSPD-----ESTPELSAEP--PKALENS 717
Db 602 YKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPSYPPTYKAKPTYKAKP 661
QY 718 PKPVGPTTKTPAA-----TKPEMTTAKDKTTERDLRTTPTTAAAP-----KMT 763
Db 662 PSYP--PTYKAKPSYPPTYKAKPSYPPTYKAKPTYKAKPTYKAKPTYKAKPSYPPTHAKPT 719
QY 764 KETATTEKTESKITATTT 783
Db 720 YKAKPTYPSTYKAKPTYPST 739

RESULT 4
US-09-103-429A-3

aal.ra1

Mon Apr 29 08:35:13 2002

Sequence 3, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
APPLICANT: Wang, Ping
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: cDNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
STREET: 118 No. 6187558th Tlloga
CITY: Ithaca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 786 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
US-09-103-429A-3
Query Match 7.6%; Score 498.5; DB 4; Length 786;
Best Local Similarity 29.1%; Pred. No. 7.1e-26;
Matches 200; Conservative 25; Mismatches 207; Indels 255; Gaps 32;
QY 156 PTPKAEITTKGPAITTPKEPTTPKPEASTTPKPEPTTIKSAPTTPKPEAPTTTKSAP 215
DB 93 PGPPAE-TTQAPATT-----QAPTTQATTTT-----QAPTTTQ-ATTTQAP 135
QY 216 TTPKPEAPTTTPKEAPTTTPKPEAPTTTKSAPTTPKPEAPTTTPKPAITTPKE 275
DB 136 TTTQ--APTTQ--ATTT---QAPTTTQ--APTTTQAPTTTQ--APTTQ--APTTQ-- 183
QY 276 PAPTTPKETPTTPKPEAPTTTPKPEAPTTAPKPAITTPKPEAPTTTPKPEAP-- 332
DB 184 -APTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 241
QY 333 -----PTTTPKPSPTTPK- 345
DB 242 ELLPNCGPADFDIHLPHDKYCNLFYQCSNGYTFEQRCPGELYPNFYVQRCDSPANVEC 301
QY 346 -----EPAPTTT----- 352
DB 302 DGEISPAPPVTEGNEDEDIDGLDLNCPANFEIDLPHGNCRDKKYQCQVHGNLVER 361
QY 353 -----KSAPTTTPKPEAPTTTKSAPTTPKPEAPTTTPKPEAPTTTPKPE 385
DB 362 CGAGTHFSFLEQCDHIELVGCTLPGGESEEDVDDBA-CTGWYCPTEPIEWEPLNPGCP 420

386 APTT-----PKEP-----APTTPKKPAPT----- 405
DB 421 ADFSIDHLLPHESDCGYLOCVHGTIARPCGNLHFSATQSCSPVTAGCQVFECDSD 480
QY 406 ---PKEAPTTTPKPEAPTTTKKPAKTAPKPEAPTTTPKETAAPTTPKLTPTTPKLAAPT 462
DB 481 NOCTSTAAPTAAPTAAPTAAPTAA--APSTVVPPTAPTAAPVPTT---AAPT--- 533
QY 463 EKPAPTTPEELAPTTPEPTTPPEEPAPTTPKAAAPT--PKEAPTTTPKPEAPTTTPKE 520
DB 534 --PAPTAAPTAAPTAAPESTTPTVTP-PTAAPTAAPTAVPEIPITV---SAPTAAPT 587
QY 521 PAPTTPKETAPTTPKGTAPTTLKPEAPTTTPKPAKELAPTTTKEPTSTTSKPAAPTTPK 580
DB 588 AAPTAAPTAAPTAVPEIPTVTSPTAAPTAA--APNT-----VTVP 632
QY 581 GTAPTTPKPEAPTTTPKGTAPTTLKPEAPTTTPKPAKELAPTTTKGPTTTS 640
DB 633 TAAPT--AAPANTTPTVPTAAPTAAPTAA--APNTAAPTVTTTS 677
QY 641 DKPAPTTPKETAPTTPKPEAPTTTPKPP 667
DB 678 ---APATTPEDDDDIDP--PLENDPINP 699
RESULT 5
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
; US-08-700-651-5
Query Match 7.4%; Score 489; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 7.8e-25;
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;
QY 29 KKNRTKKKTPKPPVWDEAGSLDNGDFKVTPTDSTTQHNKYSTSPKITTAKPINPRES 88
DB 116 RSNETK---TTEPSANTYAGVRSN-ETKTPEPSANT---NFLVDPKI-----N 158
QY 89 LPNNSDTSKE-----TSLVYNKETTVEETKETTNTKOTSTDGKETTSAK 133
DB 159 APCNSENSEFQGIIDMGSKVYIPYTKCVGVKHTTTTTTTTTTTTTTTTTTTTTTTT 208
QY 134 ETQSIENTSAKDLAPTSKVLAKTPKAEITTKGPAITTPKPEPTTPKPEASTTPKPEPTP 193
DB 209 -----TTT 240
QY 194 TTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPAITTKSA 253
DB 241 TT 300
QY 254 PTTTPKPEAPTTTPKPAITTPKPEAPTTTPKPEPTTPKPEAPTTTPKPEAPTTTPKPEAPTAPK 313

```
Db 301 TTTTTPKPTTTT----- 350
QY 314 KPAPPTPKPAPPTPKPAPPTPKPAPPTPKSAPTTKSPAPTTKAPPTTKSAPT 373
Db 351 TTTT----- 406
QY 374 PKE-----PSTT----- 381
Db 407 TSETSVIKPDEMCWLEKNGCEAKGATYVGVIGKDGRIENGMAFTMIPNDDTIVRFRFK 466
QY 382 -----TKBP-----APTPE 392
Db 467 VKDVGNTISVRCRKGAGKLEFPDRSLDFTIPPVAGHNSCSIIVGSGDKLHVSPYSGKD 526
QY 393 PAPTTPKKPAPTPKE-----PAPTTPKEPAPTTTKKAPAPKAPA 434
Db 527 VSLIS-----APIQSELENEVYCDCTAKYGAIHSGVQTSADFVTTTAKPTTT--TGA 580
QY 435 PTPPKETAPTPKKLTPTTPPEKLAPTPPEKAPPTPEELAPTPPEEPTPTTPEPAPTTP 494
Db 581 PQQPTTTTGSPPKPTTTTAKATTT-----TTLNPIITTTQKPTTT--TT- 627
QY 495 KAAAPNTEKAPAPTPKEPAPTTKEPAPTPKETAPTTPKGTAPTTLKAPAPTPPKKA 554
Db 628 --KVPGRPIATTTTTLKPIVTTTAKATTTTITVPTT-----TTTTKRDGMTTTTTL 680
QY 555 PK-----ELAPTTTKEPTS-----TTSKAPAPTPPKG--TAPTTPKEPAPT 594
Db 681 PDIGDIEITPIEKMLDKYTRMIYDYNGLLDSNDPEIPGSOAGQIADTSLNLPVQTH 740
QY 595 PKEPAPTPKGTAPTTKE-----PAPTTPKKPAPKELAPTTTTPKPTSTTSKAPAPTPKET 651
Db 741 KSTGLPDMVGLPDPKSNLHVPTNOTMGLSVSLAAKNLTVDTDETYG--LPIDT 798
QY 652 APTTPKEPAPTPKKAPAPTPPEPTTSEVSTPT-----TTKEPTTIHKS----- 697
Db 799 LTGYPLDPVSLIPN--PETGELFDPDISDEIMNGTIGAGIVSGISASESLISQKSLIDPA 856
QY 698 -----PDEST-----PELSAETPKALENSPK 719
Db 857 TNMVVGFGLLNATGVMIPGLPSPSEQTFPEIEDGGIIPPEVAANADKFKLSIP- 915
QY 720 EPGVPTTKTAAKPEMTTAKDKTTER-----DLRTTPTTTAAPKM 762
Db 916 -PSP-----ESIP-----KQKIDISSELMYDIESGRLIQVQSKRPIGSIAGDLNP 963
QY 763 TKTATTTKTTESKITATT--TQVTSSTTQDTPPKITTLKTTTLAPKVTITTKTITTE 821
Db 964 IMKTPTQDTSVTKPIDPTTGLPFPNPTGHLNPTNNNTWDSFAGAYKIYVNSGIKTDN 1023
QY 822 IMNKP--EETAKPKD-----RATNSKATTPKPKPTKAPKPTSTKKPKTMP----- 866
Db 1024 VYGLPVEITGLPKPGSDIPENSTTSELVDPSTGKPINNSTAGIVSGKPLPIEDENG 1083
QY 867 -----RVRKPTTPPKMTS--TMPE----- 886
Db 1084 NLFDPSTNLPIDGNGLVNPENSTVSGTSTTKPKGIPVNGGVVPDEAKDQADKG 1143
QY 887 -----LNPSRIAEMLOTTTTPNTPNSKLVEVPKSEDAGGAGETPHMLLRPHVEMP 941
Db 1144 KQGLIAPPTNSINKDPVTNTQVSNNTGNI-----INP--ETGKVIPLGSLPGLSPFNWTP 1197
QY 942 EVTPD-----MDYLPVRPNOGLIINP-----MLSDETNICNGK----- 974
Db 1198 QQTDEITGKPVDTVGLPVPDSTGEIIDPATKLPVGSVAGDEILTEVLNITTDVETGLP 1257
QY 975 -----PVDGLTTLNGLTFLV 988
Db 1258 IDLETGLPRDPVSGLPQLPNGTLV 1281
```

RESULT 6

```
US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, HANA
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1721 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-928-361B-6
```

```
Query Match 7.4%; Score 489; DB 3; Length 1721;
Best Local Similarity 22.1%; Pred. No. 7.8e-25;
Matches 284; Conservative 82; Mismatches 476; Indels 442; Gaps 41;

QY 29 KKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTSTOHNKYSTSPKITTAKPINRPS 88
Db 116 RSNETK---TTEPSANTYAGVYRSN-ETKTEPSANT---NFLVDPKI-----N 158
QY 89 LPPNSDTSKE-----TSLTVNKETTVETKETTTTNKQSTDGEKETSATK 133
Db 159 APCNSENSEFQGGIFDMGSKVYIPYTKCVGKRHTTTTTTTTTTTTTTTT----- 208
QY 134 ETOSIEKTSADLAPTSTKVLAKPTPKAETTKGPAITTPKEPTTPKPEASTTPKEPTP 193
Db 209 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 240
QY 194 TTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKKEPAPTTTPKEPAPTTTKSA 253
Db 241 TTTT----- 300
QY 254 PTTTPKEPAPTPPKKAPPTTPKPEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 313
Db 301 TTTT----- 350
QY 314 KPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSPAPTTTKSAPTTTKSAPT 373
Db 351 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 406
```



```

QY 655 TPKEP-----APTTPKKAPATTPTTSEVSTPTTKPTTIHKSPDESTPELSAE 708
Db 650 LPIIPQKGNIPNLNPNSEVSEYPRNDNGENSNMTKSKKNI---PNEPIPSPGDN 706
QY 709 PTPKALENSPK-----EPGVPTTK 727
Db 707 PYKGHEERPKPHRSNDYVYDNNVNNKNDPEIPNNE 745

RESULT 8
PCT-US92-00018-2
; Sequence 2, Application PC/TUS9200018
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hedstrom, Richard
; APPLICANT: Khumsmith, Srisin
; APPLICANT: Rogers IV, William O.
; TITLE OF INVENTION: Protective malaria sporozoite surface protein
; TITLE OF INVENTION: Immunogen and gene encoding
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: A. David Spevack
; STREET: NMDC Building 1 T-12 National Naval
; CITY: Medical Center
; STATE: Bethesda
; COUNTRY: USA
; ZIP: 20814-5044
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/00018
; FILING DATE: 19920103
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, Avram D.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-4033
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-00018-2

```

```

Query Match 7.4%; Score 488.5; DB 5; Length 826;
Best Local Similarity 26.4%; Pred. No. 3.5e-25;
Matches 153; Conservative 54; Mismatches 223; Indels 149; Gaps 34;

QY 192 TPTTIKSAPTTKEPA-PTTTKSAPTTKEPA-PTTTKEP-APTTPKEP-APTTPKE 244
Db 273 TPCKVRDCPOIPPPVPIPNKIPEKPSNPPEVNPDPNDPNNPNNPNNPNNPNN 332
QY 245 PAPTTPKSAPTTKEP-APTTPKKP-APTTPKEP-APTTPKEP-APTTPKEP 300
Db 333 PNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNNPNN 392
QY 301 PPTTPKEPAPTAPK--KPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 358
Db 393 RRPKRNNPKPKNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKNPKN 443
QY 359 TKEPAPTTPKSAPTTKEPSPPTTKPEAPTTPKEPA-PTTPKKP-APTTPKEPAPT 416
Db 444 PNKPNNP-----NEPSNKNPNNP-----EPLNPNPNNPNNPNNPNNP 492
QY 417 PAPTTPKKPAPTAPKEP-APTTPKETAPTTPKKLTPTTPEKAPTTPKEPAPT 475

```

```

Db 493 PS-----NPNEPSNPNEPSNPNE-----PSNPKK-----PSNPNP--P 523
QY 476 TTPPEPTTPPEAPATTTPKAAAPNTPKP-APTTPKEPAPTTPKBPAPTTPKETAPT 534
Db 524 SNPNE--PLAPNEP-----SNPNEPSNPNEPSNPNE--PSNPKB--PSNPNP 564
QY 535 KGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDKPAPTTPKGTAAPTTPKE 594
Db 565 -----PSNPEEPNPEE--PSNKEP-----SNPEEPINPEELNPKPESNPEES 606
QY 595 PKEPAPTTPKGTAAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSDKPAPTTPKE 654
Db 607 PKEPI-----NPEESNPKP-----INPEDNENPLIODEPIEPNDSNVPI 649
QY 655 TPKEP-----APTTPKKAPATTPTTSEVSTPTTKPTTIHKSPDESTPELSAE 708
Db 650 LPIIPQKGNIPNLNPNSEVSEYPRNDNGENSNMTKSKKNI---PNEPIPSPGDN 706
QY 709 PTPKALENSPK-----EPGVPTTK 727
Db 707 PYKGHEERPKPHRSNDYVYDNNVNNKNDPEIPNNE 745

RESULT 9
5202236-13
; Patent No. 5202236
; APPLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLISS, RUSS; WEI, TENA; FILPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIONADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456
; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO: 13
; LENGTH: 652
5202236-13

```

```

Query Match 7.3%; Score 476.5; DB 6; Length 652;
Best Local Similarity 29.7%; Pred. No. 1.7e-24;
Matches 221; Conservative 71; Mismatches 284; Indels 169; Gaps 45;

QY 75 PKITAKPINPRPSLPPNSDTSKETSLSLVNKEITVETKETTNNKQTSOGKEKTTSAKE 134
Db 11 PKMTYPTTKPKPSYPP-----TYKSKPTY-----KPKIT----- 40
QY 135 TQSIKTSADLAPTSKVLAKPTKPAETTTKGPAALTTPKEPTTTPKE-----PASTTPKE 190
Db 41 -----YPTTK--AKPS-----YPTTKPKTYPTTKPKLTYPPTTKPKP 79
QY 191 PTPPTIKSAPT--TPKEPAPTTPKSAPTTTPKEPAPTTPTTKEPAPTTPTTKEPA--P 247
Db 80 SYPTTKSKPTKPKTYKTYPTTKPKPSYPTTKPKKTYPTTKPKKTYPTTKPKASYP 138
QY 248 TTTKSAPTTTPKEAPTTPKKAPAT--TPKEPAPT-----PKEPTTPKEPAPTTPKEAPT 302
Db 139 PTKPKPSYP--PSYTKTKTYPTTKPKLTYPPTTKPKPSYPTTKPKTYPTTKPKLT 196
QY 303 TPKEPAPTAPKPA--PTTPKEPA--PTTPKEPAPTTPKEPSPTTP-----KEPAPTTPK 353
Db 197 YP-----PTYKAKPSYPTTKPKPSYPTTKPKPSYPTTKPKPSYPTTKPKPSYPTTK 252
QY 354 SAPTTTKEPA-----PTTTKSAPTTTPKEPSPTTPKEPA--PTTPKEPA-----PTTPKK 400
Db 253 AKPTYKAKPTYKAKPTYKAKPTYKAKPTYKAKPTYKAKPTYKAKPTYKAKPTYKAK 308

```

[illegible]

```

: APPLICANT: WIND, MARC W.T.
: APPLICANT: WIND, RICHÉLE D.
: APPLICANT: VAN DEN BOSCH, TANJA J.
: TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
: TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
: TITLE OF INVENTION: PREPARATION THEREOF
: FILE REFERENCE: 2728-2
: CURRENT APPLICATION NUMBER: US/09/219,849
: CURRENT FILING DATE: 1998-12-23
: NUMBER OF SEQ ID NOS: 50
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 5
: LENGTH: 960
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Illustrative
: OTHER INFORMATION: amino acid sequence
: US-09-219-849-5

Query Match          6.9%; Score 452; DB 4; Length 960;
Best Local Similarity 25.6%; Pred. No. 1.2e-22; Indels 134; Gaps
Matches 173; Conservative 60; Mismatches 309;

QY 156 PTPKAETTTKGFALTTPKEPTPTTPKEPASTTPKEFTPTTIKSAPTTKPEAP--TTTTS 213
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 PGPAGPGPSRDGP--PPCAFGPAGP--PGSRDGPFGAGP-PAGPPGSRDGPFGAPGPA 126

QY 214 APTTKEPAPTTTKEPA-PTTTPKEPAPTTTKEPATTTKSAPTTTKEPAPTTT 369
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 GPPGSRDGPFGPAGPAGPAGPSRDGPFGAGPAGPSRDGPFGAGPAGPAGPSRDGP 186

```


aal.ra1

Mon Apr 29 08:35:13 2002

STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,116
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning 128 to 1727 constitute a repeated region wherein the repeat is 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA or ACG; and Asn = AAT or AAC."
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA or CCG; and Ala = GCT, GCC, GCA, or GCG."
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a 21 amino acid precursor sequence."
OTHER INFORMATION: 21 amino acid precursor sequence."

Query Match 6.5%; Score 424.5; DB 4; Length 1867;
Best Local Similarity 23.7%; Pred. No. 1.8e-20;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;
Qy 99 TSLVTKETTVETKETTNNKQTSKGTSTGKETTSAKETOSIEKTSKAD-LAPTSTKVLAKPT 157
Db 16 TVLIV-----VTGSGHASSTPGGKETTSATORSVSPSTKNAVMSMTSVLSSHS 65
Qy 158 P-KAETTTKGP--ALTTPKEP-----TPTTPKEPASTTPKPTPTTTIKSAP 200
Db 66 PGSGSSTTQGDVTLAPATEPASGAATWGDVTSVPTPALGSTTPPAHDVT---SAP 122
Qy 201 TTPKEPAPTPT-----KSAPTTPKEP-----APTTPKEPAPTTPKEPAPT 241
Db 123 --DNKPAGSTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 180
Qy 242 TKE--PAPTPT-----KSAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPKEPAPT 288
Db 181 APDXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 240
Qy 289 PKE--PAPTTPKEP-----APTTPKEPAPTTPK---KPAPTTPKEPAPTTPKEPAPT 335
Db 241 APDXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 300
Qy 336 TKEPSPTTPKEPAPT--TKSAPTTPKEPAPT-----TKSAPTTPKEPAPTTPK---E 384
Db 301 APDXRP--XPGSTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 359
Qy 385 PAPTTPKEPAPTTPK---KPAPTTPKEPAPTTPK---EPAPTTPKKEPAPTAPKEPAPT 436
Db 360 SAPDXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 419
Qy 437 TPKETAP-----TTPKKLPTTPEKLAPTTPEKAPAPTTPKEPAPTTPKEPAPTTPKGT--APTTLKEPAP 547
Db 420 SAPDXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 492
Qy 493 TPKA---AAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKGT--APTTLKEPAP 547
Db 472 APXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 530
Qy 548 TTP-----KKPAPKELAPTTTKEPTSTTSOKPAP--TTPKGTAPTTPKEPAPTTP 595
Db 531 TAPXAHGVTSPADXRPPGSGTAP--XAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 585
Qy 596 KEAPTTTPKG---TAPTTLKEPAPTTP-----KKPAPKELAPTTTKEPTSTTS 641
Db 586 PXPFGSTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 644
Qy 642 KPAP--TTPK-----ETAPTTPKEPAPTTP-----KKPAPTTPPETTSEVSTP 685
Db 645 RPXPGSTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 702
Qy 686 TTTKEPT-----IHKSPESTPELSAEPPTPKALENSPKPEGVPTTTPAA----- 731
Db 703 DXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 762
Qy 732 -TKPEMTTAKDK---TTERDLRTTPTTTAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 787
Db 763 DXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 820
Qy 788 TTTQDPTTTPFKITTLKTTTLAPKVTITTKTITTTIMNKPEETAAPKDRATNSKATTPKPK 847
Db 821 APDXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 868
Qy 848 KPTKAPKPTSTKPKMTKPRVRKPTTTPKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMTSTMPKMTSTMP 907
Db 869 -GSTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 911
Qy 908 -PNSKLVENVNPKSEDAEGETPHMLLRPHVFMPEVTPDMDYLPVFNNOGIIINPM 966
Db 912 APXAHGVTSPADXRPPGSGTAPXAH-----GVTSAPDXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVTSPADXRPPGSGTAPXAHGVT 957
Qy 967 ETNICNGKPVDTLTT--LRNGTLVAFRGHYFWMLSPFPSPARRITEVWGPSPID 1021

Db 958 VTSAPDXRXPXPGSTAPXAHGVTSA-----PDXRXPXPGSTAPXAHGVTSPAD 1003

RESULT 13

US-08-479-537A-2

Sequence 2, Application US/08479537A

Patent No. 5861381

GENERAL INFORMATION:

APPLICANT: CHAMON, Pierre

APPLICANT: KIENY, Marie-Paule

APPLICANT: LATHE, Richard

APPLICANT: HAREUVENT, Mara

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,537A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 90/13101

FILING DATE: 23-OCT-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 017753-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-6621

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 128..1899

OTHER INFORMATION: /note= "The amino acids spanning

OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat

OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such

OTHER INFORMATION: repeats varies from 1 to 40."

FEATURE:

NAME/KEY: Peptide

LOCATION: 134

OTHER INFORMATION: /note= "Amino acid 134 is X1 = xaa

OTHER INFORMATION: Xaa Xaa which is the codon for pro or Ala wherein pro = CCT,

OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 144

OTHER INFORMATION: /note= "Amino acid 144 is Y = xaa

OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,

OTHER INFORMATION: or ACG; and Asn = AAT or AAC."

FEATURE:

NAME/KEY: Peptide

LOCATION: 147

OTHER INFORMATION: /note= "Amino acid 147 is X2 = xaa

OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,

OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..21

OTHER INFORMATION: /note= "Amino acids 1 to 21 are a

OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2

Query Match 6.5%; Score 424.5; DB 2; Length 2035;

Best Local Similarity 23.7%; Pred. No. 2e-20;

Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 99 TSLVNTKETTETKTTTNTKOTSTDGKEKTTSAKETQSIKTSKD-LAPTSTKVLAKPT 157

Db 16 TVLTV-----VTGSHASSTPGGKETSATQSSVPSSTEKNAVMTSSVLSHS 65

QY 158 P-KAETTTKGP--ALTTTPKEP-----TPTTPKEPASTTTPKEPTTIKSAP 200

Db 66 PGSGSTTGGQDVTLAPATEPASGSAATWGQDVTSVPVTRPALGSTTTPAHDTV---SAP 122

QY 201 TTPKEPAPTTT-----KSAPTTTPKEP-----APTTPKEPAPTTTPKEPAPT 241

Db 123 ---DNKPAPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 180

QY 242 TKE--PAPTTT-----KSAPTTTPKEPAPTTPK---KPAPTTTPKEPAPTTTPKEPTT 288

Db 181 APDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 240

QY 289 PKE--PAPTTKEP-----APTTPKEPAPTPAK---KPAPTTTPKEPAPTTTPKEPAPT 335

Db 241 APDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 300

QY 336 TKEPSPTTPKEPAPTT--TKSAPTTTTPKEPAPTT-----TKSAPTTTPKEPSPTTTPK---E 384

Db 301 APDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 359

QY 385 PAPTTTPKEPAPTTPK-----KPAPTTTPKEPAPTTPK-----EPAPTTTTPKEPAPTT 436

Db 360 SAPDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 419

QY 437 TPKEPAP---TTPKKLTPTTPKEPAPTTTPKEPAPTTPEELAPTTPEPTTTPKEPAPT 492

Db 420 SAPDXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 471

QY 493 TPKA-----AAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKGT--APTTLKEPAP 547

Db 472 APXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 530

QY 548 TTP-----KKPAPKELAPTTTKEPTSTTSDKPAP---TTPKGATPAPTTTPKEPAPTT 595

Db 531 TAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 585

QY 596 KEAPTTTPKG---TAPTTLKEPAPTT-----KKPAPKELAPTTTKEPTSTTSD 641

Db 586 PXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 644

QY 642 KPAP--TTPK-----ETAPTTTPKEPAPTT-----KKPAPTTTPKEPTTTPKEPTT 685

Db 645 RPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 702

QY 686 TTTKEPTT-----IHKSPDESTPELSAETTPKALENSPKPEPGVPTTKTAA----- 731

Db 703 DXRXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVT 762

QY 732 -TKPEMTTAKDK---TTERDLTTTETTAAPKMKETATTEKTESKIATTTQVTS 787
Db 763 DXRPFSTAPXAHGVTSPDXRPFST--APXAHGVTSPDXRPFSTAPXAHGVT 820
QY 788 TTTQDTPFKITLTKTTTLLAPKVTITTKITTTTEIMNKPEETAAPKDRATNSKATPKPQ 847
Db 821 APDXRPFST---STAPXAHGVT-----SAPDXRPFSTAPXAHGVTSPDXRPFST- 868
QY 848 KPTKAPKKTSTKPKTMTPRVRKPKTTPTRKMTSTMPENLPTSRIAEAMLOTTPTRPNOT 907
Db 869 -GSTAPXAHGVTSPDXRPFST--XPGSTAPXAHGVTSA-PDXRPFST-----PGST 911
QY 908 -PNSKLVENPKSEDAGGAGETPHMLLRPHVEMPEVTPDMOYLPRVPGOGIIINMLSD 966
Db 912 APXAHGVTSPDXRPFSTAPXAHGVT-----GVTSPDXRPFSTAPXAHGVTSA-----PXAHG 957
QY 967 ETNICKPKVDGLTT-LRNGTLVAFRGHVFWMLSPFSPSPARRITEVVMGIPSPID 1021
Db 958 VTSAPDXRPFSTAPXAHGVTSA-----PDXRPFSTAPXAHGVTSPDXRPFSTAPD 1003

RESULT 14
US-09-083-116-2
: Sequence 2, Application US/09083116
: Patent No. 6203795
: GENERAL INFORMATION:
: APPLICANT: CHAMBER, Pierre
: APPLICANT: KIENY, Marie-Paule
: APPLICANT: LATHE, Richard
: APPLICANT: HAREUENI, Mara
: TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
: TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM: disk
: MEDIUM TYPE: IBM PC compatible
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/083,116
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/479,537
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/FR91/00835
: FILING DATE: 23-OCT-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/039,320
: FILING DATE: 04-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/403,576
: FILING DATE: 14-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Teskin, Robin L.
: REGISTRATION NUMBER: 35,030
: REFERENCE/DOCKET NUMBER: 017753-025
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2035 amino acids
: TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
128 to 1899 constitute a repeated region wherein the repeat
of 20 amino acids, 17 of which are fixed. The number of such
repeats varies from 1 to 40."
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 = Xaa
Xaa Xaa which is the codon for Pro or Ala wherein Pro = CC
CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y = Xaa
which is the codon for Thr or Asn wherein Thr = ACT, ACC,
or ACG; and Asn = AAT or AAC."
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 = Xaa
which is the codon for Pro or Ala wherein Pro = CCT, CCC,
or CCG; and Ala = GCT, GCC, GCA, or GCG."
OTHER INFORMATION:
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
21 amino acid precursor sequence."
OTHER INFORMATION:
US-09-083-116-2

Query Match 6.5%; Score 424.5; DB 4; Length 2035;
Best Local Similarity 23.7%; Pred. No. 2e-20;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;
QY 99 TSLTVNKETTETKETTNTKOTSTDKKETTSAKETQSIKTSKAD-LAPTISKVLAKPT 157
Db 16 TVLTV-----VTGSHASSTPGGKETSATQSSVPSSTKNAVSMISVLSHS 65
QY 158 P-KAETTTKGP--ALTPKPE-----TPPTKPEASTTPKPEPTTTKSP 200
Db 66 PGSGSTTQGDVTLAPATEPASGSAATWGQDVTSPVTRPALGSTTPPAHDVT---SAP 122
QY 201 TTPKPEAPTTT-----KSAPTTKPE-----APTTTKEPAPTTPKPEAPT 241
Db 123 --DNKPAGSTAPXAHGVTSPDXRPFSTAPXAHGVTSPDXRPFSTAPXAHGVT 180
QY 242 TKE--PAPTTT-----KSAPTTKPEAPTTPK-----KPAPTTPKPEAPTTPKPT 288
Db 181 APDXRPFSTAPXAHGVTSPDXRPFSTAPXAHGVTSPDXRPFSTAPXAHGVT 240
QY 289 PKE--PAPTTKPE-----APTTKPEAPTAPK-----KPAPTTPKPEAPTTPKPEAPT 335
Db 241 APDXRPFSTAPXAHGVTSPDXRPFSTAPXAHGVTSPDXRPFSTAPXAHGVT 300
QY 336 TKEPSTTPKPEAPT--TKSAPTTTKEPAPT-----TKSAPTTTKEPSTPTTK----E 384
Db 301 APDXRPFSTAPXAHGVTSPDXRPFSTAPXAHGVTSPDXRPFSTAPXAHGVT 359
QY 385 PAPTTTPKPEAPTTPK-----KPAPTTPKPEAPTTPK-----EPAPTTPKPAKPEAPT 436
Db 360 SAPDXRPFSTAPXAHGVTSPDXRPFSTAPXAHGVTSPDXRPFSTAPXAHGVT 419
QY 437 TPKETAP-----TTPKKLTTPTPKLPAPTTPKPEAPTPEELAPTTPEPTTPPEAPT 492
Db 420 SAPDXRPFSTAPXAHGVTSPDXRPFSTAPXAHGVTSPDXRPFSTAPXAHGVT 471
QY 493 TPKA---AAPNTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 547

[illegible]

RESULT 15

US-08-276-967-2
; Sequence 2, Application US/08276967
; Patent No. 5851817
; GENERAL INFORMATION:
; APPLICANT: Hardy, Daniel M.
; APPLICANT: Garbers, David L.
; TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of
; TITLE OF INVENTION: Sperm
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,967
; FILING DATE: Submitted Herewith
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: UTSD:418\KIT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679

```

:      TELEX: 79-0924
:
:      INFORMATION FOR SEQ ID NO: 2:
:
:      SEQUENCE CHARACTERISTICS:
:
:      LENGTH: 2476 amino acids
:
:      TYPE: amino acid
:
:      STRANDEDNESS: single
:
:      TOPOLOGY: linear
:
:      MOLECULE TYPE: protein
:
:      US-08-276-967-2

```

Query Match	6.4%	Score 419.5;	DB 2;	Length 2476;
Best Local Similarity	32.5%;	Pred. No.5.5e+20;		
Matches 163;	Conservative 43;	Mismatches 177;	Indels 119;	Gaps 31;
QY	430	PKEPAPPTPKETATPTPKLTPITPEKALATPTPEKAPPTPEELATPTDEPTPTTPEP	489	
Db	313	PSETSVSEKPVAPTE---KPTVPSEIYITPEKPMVMHMEKPIVHT--EKPTVPT-EKP	365	
QY	490	APTTPKAAAPNTKPEPA-----PTTPKPEAPPTPKPEAPPTPKETATPTPKGAPITLKEP	545	
Db	366	TIPTKEKSVPT--KKPVFEKPTLPPE-GPTVPAAE-RPTTPPEGVAPPKG--PTVLTE-	418	
QY	546	APPTPKDAPKELAPPTTKBETSTSDKAPPTTPKGTAPT--TPKEPAPPTPKPEAPPTPK	604	
Db	419	-----WPTSHTEKSTVHTEKPILTGKSTIPEKPMVPTKRT-----TTP-	458	
QY	605	GTAPTTLKEAPPTPKKPAKELAPT--TKGPTSTT---SDKAPPTPKETATPTPKPEPA	660	
Db	459	-TERTTPAEKPTV---PIEKPMVPTERTIPTERTIPTERTIPTERTIPTERTIPTERTIPT	512	
QY	661	PTTPKKAPPTTPEPPP-----TTSEVSTPT-----TTKEPTTIHKSDESTPELSAPT	710	
Db	513	VPTEKPIVPTKEHTIPTPEKLTVLTEHTIPTERTIPTERTIPTERTIPTERTIPTERTIPT	571	
QY	711	PKALESKPEGVPTTKPAATKPEMTTAKDKTTERDLRTTPTETTTAAPKMTETATTT	770	
Db	572	-----EETPTPEKLTVPST--ERTTTPKTRTTPRTIRTTPIT-----RTTPT	613	
QY	771	EKTTESKITATTTQVSTTTQDPTTPKTIITLKTITLAPKVTITTKTITTTIMKNKPETA	830	
Db	614	ERTT-----TPIRITPTPERTT---IPTKKTT-----VPTEKTIITPT-----ERTI	652	
QY	831	KPKDRATNSKATTPKQKPTKAPKPKPTSKPKPTMPRVKPKPTTTPPKMTSTMPENLPT	890	
Db	653	AP-----TTPQP--SETLVPQPAVMPST-----SATVTPTPTTIASCP---PN	693	
QY	891	SRNAEAMLTQTTTRPNOTPNSKL	912	
Db	694	AHFERCACPVSCQ--SPTPNCEL	714	

Search completed: April 26, 2002, 16:25:14
Job time: 454 sec

aal.raai

Mon Apr 29 08:35:13 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:30:58 ; Search time 114.61 Seconds
(without alignments)
905.906 Million cell updates/sec

Title: AA2
Perfect score: 7276
Sequence: 1 MAWKLPYIYLLLSVFVIQ.....ARAITRSGQLSKVWVNC 1363

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1077	14.8	3020	2 A43932	mucin 2 precursor,
2	950	13.1	1664	2 T18262	S-layer protein -
3	863.5	11.9	1489	2 T31108	cyst germination s
4	846	11.6	1274	2 T16251	hypothetical prote
5	812	11.2	2187	2 T30826	nascent polypeptid
6	809	11.1	1367	1 S48478	glucan 1,4-alpha-g
7	784	10.8	1188	2 S49915	extensin-like prot
8	695	9.6	1151	2 T18535	high molecular mas
9	684.5	9.4	1229	2 T25697	hypothetical prote
10	672	9.2	1344	1 A35175	mucin 1 precursor,
11	666.5	9.2	7962	2 T38346	elastic titin - hu
12	653	9.0	3570	2 T45025	mucin MUC5B, trach
13	643.5	8.8	990	2 T51618	nucleolar phosphop
14	636	8.7	3507	2 T34513	hypothetical prote
15	633	8.7	489	2 T11622	extensin class 1 p
16	632	8.7	761	2 C84672	hypothetical prote
17	629	8.6	6642	2 T29757	protein UNC-89 - C
18	627.5	8.6	971	2 T19431	hypothetical prote
19	622.5	8.6	839	2 T75518	hypothetical prote
20	607.5	8.3	801	2 T29018	hypothetical prote
21	607.5	8.3	924	2 S27923	gene LF3 protein -
22	605	8.3	378	2 S50125	larval glue protei
23	591.5	8.1	2232	2 T34434	hypothetical prote
24	588.5	7.8	1630	2 A53577	ascites sialoglyco
25	559.5	7.7	350	2 S22456	hydroxyproline-ric
26	559	7.7	875	2 S23760	polyphenolic adhes
27	556.5	7.6	856	2 T16543	hypothetical prote
28	551	7.6	620	2 S06733	hydroxyproline-ric
29	551	7.6	1087	1 QFM5H	neurofilament trip

ALIGNMENTS

RESULT 1
A43932
mucin_2 precursor, intestinal - human (fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C:Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C:Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329
J:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J: Biol. Chem. 269, 2440-2446, 1994
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A:Reference number: A49963; MUID:94132002
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
A:Cross-references: GB:L21998
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A:Reference number: A45106; MUID:93016075
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:gl86395; PIDN:AAA59163.1; PID:gl86396
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:gl86397; PIDN:AAA59164.1; PID:gl86398
A:Experimental source: colon
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A:Reference number: A43932; MUID:9135817
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:gl88863; PIDN:AAA59875.1; PID:gl88864
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A:Reference number: A43532; MUID:89197956
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-references: GB:M22405; NID:gl88873; PIDN:AAA36334.1; PID:gl88874
A:Experimental source: intestine
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

calphotin - fruit
hydroxyproline-ric
hypothetical prote
extensin precursor
neurofilament trip
hypothetical prote
vsaA protein precu
exo-alpha-stallidas
calcium-binding pr
hydroxyproline-ric
membrane glycoprot
hypothetical trip
hypothetical prote
tagument protein 2
mucin-like glycopr
membrane glycoprot

J. Clin. Invest. 87, 77-82, 1991

A:Title: Human bronchus and intestine express the same mucin gene.

A:Reference number: A61257; MUID:91086481

A:Accession: A61257

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>

A:Experimental source: bronchus

R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstner,

Biochem. Biophys. Res. Commun. 183, 821-828, 1992

A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-

A:Reference number: P00328; MUID:92198477

A:Accession: P00328

A:Molecule type: mRNA

A:Residues: 2328-2468 <XUG>

A:Cross-references: GB:M8523

A:Experimental source: small intestine

A:Accession: P00329

A:Molecule type: protein

A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>

C:Genetics:

A:Gene: GDB:MUC2

A:Cross-references: GDB:120203; OMIM:158370

A:Map position: lip15.5-lip15.5

C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von

C:Keywords: glycoprotein; intestine; tandem repeat

F:2766-2834/domain: von Willebrand factor type C repeat homology <VWC>

Query Match 14.8%; Score 1077; DB 2; Length 3020;
Best Local Similarity 27.2%; Pred. No. 1.4e-37;
Matches 373; Conservative 103; Mismatches 497; Indels 398; Gaps 50;

QY 36 SFERGRDC-----DAQCKYDK-----CPVESCAEVHNPSPSKKAP 79
DB 1081 SCDTGGDECFCSAVAYAECKEGACVFWRTDLC-----IFC-DYINPHECEWHYEP 1136
QY 80 PPSASQTIKST-----KSPK--PNKKKKKVIKIESE-----ITEEH- 117
DB 1137 CGNSKFCTRTINGIHNSVYLEGYPCPKDRIYEDLKKVYADKCGVEDTHY 1196
QY 118 ---SVSENOESSSSSSSTIWKSSKNSAANRELKLVKDNKNKRNKKKPTPK 174
DB 1197 PPGASVPEETCKSCVCTNSQVCRPEEG-----ILNQTQ----- 1233
QY 175 PPVDEAGSLDNGDF---KVITPDTSTOH---NKVSTPKIPTAKPINRPSLPPNSD 228
DB 1234 -----DGAFCWEICGPNGVKEHFNICSITRPSILTTITTLTTP----- 1277
QY 229 TSKESTLVNKEITVETKET---TTTNK-----QTSQDGEKSTTS---AKET 269
DB 1278 ---TSFTITTTTPTSSVLTSTPKLCLLSDWINEHPSSGSDGDRPFDGVCAP 1333
QY 270 OSIETSAKD----- 279
DB 1334 EDIECRSVKDPHLSDHQKQVQCDVSVFTCKNEQDQGNPGLCYDYKIRVNCQPM 1393
QY 280 --LAPTSVKLAKPTKAEITTKGPAITTPKEPTTPKEPASTTPKEPTTIKSA- 336
DB 1394 KCIITPSPPTTPSPPTTTTLPTTPSPPTTTTPPTTPPTTPPTTPPTTPPTTT 1453
QY 337 PKEPATTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 395
DB 1454 PSPPISTTTTPPTTPSP-PPTTSPPTTTPSPPT-TTTPPTTPPTTPSPPTTPPT 1511
QY 396 PTPPKPAPTTKPEAPTTKPEPTTPKEAPTTKPEAPTTKPEAPTTKPEAPTTK 455
DB 1512 ASTTLPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPT 1570
QY 456 EPAPTTKPEAPTTKPEPTTPKPEAPTTKSA-TPPKPEAPTTKSA-TPPKESP 513
DB 1571 TTPPTTPSPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPT 1630

QY 514 TTTKEPAPTTKPEAPTTKPKAPTTKPEAPTTKPEAPTTKPKAPTTKPKAPTTK 573
DB 1631 STTTLPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPTTPPT 1689
QY 574 ETAPTTKLTTPTEKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPE 633
DB 1690 -SPPTT---TMTPTS---PTTTPSSPTTTTTPSSPTT-SPPTTMTTTPSPPT 1741
QY 634 NTPKEPAPTTKPEAPTTKPEAPTTT---KETAPTTT-----KG----- 670
DB 1742 TMTTLPPTTSSPTTTLPLPSITPTTTPSFSTTTTTCVPLCNWGLWDSKPNFHK 1801
QY 671 ----- 670
DB 1802 PGGTELGIDVCGPGWAANISCRATMYPDVPIGOLGTVVCDVSVGLICKNEQKPGVI 1861
QY 671 -----TAPTTLKEPAPTTKPKAPKELAPTT--TKETSTSTSKP-- 708
DB 1862 PMAFCLANYEINVOCCECVQPTTM---TTTTEMTPTTPITTTTTPPTTTPSTQSPNG 1918
QY 709 --APT--TPKGTAPTTPKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPKAPKELAPTT 765
DB 1919 LQAPPTPISITTTTTPPTTGTGTPT--TTPITTTTTPPTTGTGTGTPTTTLITTT 1977
QY 766 TKGPTST--TSKAPAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTSEVS 817
DB 1978 TMTPTPTSTKSTTTPITTTTTPPTTGTGTGTPTTGTGTGTGTGTGTGTGTGTGT 2037
QY 818 TPTTTK-----EPTTIHKSPDESTPELSAEPKALENSKPEKPVTTTKTTPAAKPEMTT 872
DB 2038 PPTHSTAPIALTTNSPPSSSTPQTSSTSSPLTESTTLLSTLPPALEMTSTAPPSTP 2097
QY 873 TAKDTERDLRTTPETTAAPKMTKATATTEKTESKITATTTQVSTTTTODTTPFKI 932
DB 2098 TAPTTTSGGHTLSPSPSTTSPPTTGTGTGT--GSSAPTPSTVOTTTTSAWTPPTPL 2155
QY 933 TT---LKTTTLAPKVTTKTITTTINMKPE----- 962
DB 2156 STPIIRTTGLRPSSVLIQVNDYAPGEVYNGYDTCYFVNCSLCTLEFYNNW 2215
QY 963 --TAKPKDRATNSKATTPKPKPTKAPKPTSTKPKKMPVRKPKKTTTPPKMTSTWPE 1020
DB 2216 SCPSTSPPTPSK--STPTSPKSPSTSPKPTGTPKPECPDFDPPR-----Q 2261
QY 1021 LNPTSIAMQLQTTTRPNQTPNSKLVEVP-----KSEDAG----- 1058
DB 2262 ENETWLCDFEM-ATCKYNTVEIVKVECEPPMPPTCSNGLQVPRVEDPDGCGWHWCDC 2320
QY 1059 ---AEGETPHMLL-----RPHVFMPEVTPDMOYLPRVNOGIIN 1095
DB 2321 YCTGWGD-PHYVTFDGLYYSQGNCTYVLVEISPSVD-----NFGVYID 2364
RESULT 2
T18262
S-layer protein - Clostridium thermocellum
C:Species: Clostridium thermocellum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18262
R:Fujino, T.; Beguin, P.; Aubert, J.P.
J. Bacteriol. 175, 1891-1899, 1993
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose.
A:Reference number: Z18847; MUID:93209931
A:Accession: T18262
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1664 <FUJ>
A:Cross-references: EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841.1

Query Match 13.1%; Score 950; DB 2; Length 1664;
Best Local Similarity 31.0%; Pred. No. 1.5e-32;

Db 1048 ASTEETTYAPTEETTYAPAEETPYEETBETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1107
QY 780 TTPKETAPTPKPPAPPTTPKPPAPPTTPPTTSEVSTPTTKBPT--TIHKSPESTP 837
Db 1108 ETTYAPAEETPYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1166
QY 838 ELSAETPKA-LENSPEKG-----VPTTKTAATKPEMTTAKDKTTERDLRTT----- 887
Db 1167 YAPTEETPYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETPYEPT 1226
QY 888 ETTTAAKMTKETATTEKTESKIITATTQVTSSTTQDPT-----PFKITT----- 934
Db 1227 EETTYAP-----TEETTYEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE 1281
QY 935 LKTTTAPKVTTHKIIITTEIMNKPEETAKPKDRAT--NSKATTPK-OKPTKAPKKT 991
Db 1282 TEETTYAPTEATTYAPTEETTYAPTEETTYEPTTETTYAPTEETTYAPTEETTYAPTEET 1341
QY 992 -----STKKP-KTMPRVKPKTTPTRKMTSTMPELNPTSRIAEMLQTTTRP 1038
Db 1342 PYEABESTSVSEKPCNTEETDEPTDEPTDE--PSDEPTDEPTDEPTDLPTDEPSTP 1399
QY 1039 --NOTPNSKLVENPKSEDAG 1057
Db 1400 CDNOGINGIGVENKVRYNAG 1420

RESULT 4

hypothetical protein F35A5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T16251
R:Leimbach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: Z18485
A:Accession: T16251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1274 <LEI>
A:Cross-references: EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB52641.1; GSPDB:GN00
A:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP:F35A5.1
A:Map position: x
A:Introns: 1272/2

Query Match 11.6%; Score 846; DB 2; Length 1274;
Best Local Similarity 27.6%; Pred. No. 2.4e-28;
Matches 313; Conservative 119; Mismatches 455; Indels 248; Gaps 60;

QY 73 PSKKAPPPSGASQTIKSTTKSP---KPP---NKKTKTKVIESEETIEHSVSENDES 125
Db 185 PSPKKAAPSKEHDPVVPPTPIKNPKAKKWPWEDDEVTEIEKEPEPATRKVPALKKKEP 244
QY 126 SSSSSSSSSSTIWLKSSKNSAANRELQKLVKVDNK-KNRTKK-KP-----TP----- 173
Db 245 STSVKPVSDPSPTKKY-----PVKKEVEVPTPIKNTKKWKPPWEDETPEEVK 294
QY 174 KPVWDEAGSLDNGD-----FKVTTPTDTSTOHKNKYSTSPKITTAKPI-NRPSLPPNSD 228
Db 295 EPPVPEKAPVLKKDPAPAAKARDSPSKAAPKKVEPSPVVPPTPVKNPVKKYKPPWE 354
QY 229 TSSETSITVYKNETVETKETTITNKOTSTDGKKTTSKETSIAKDSIAKDLAPTSKVL 288
Db 355 VDDEPAEVEVKPSAPKTKTVPVKRKEPESSSTTPSSDPSKKAAPAVKPRDSSPKATPL 414
QY 289 KPTPKAP-----TTTKGPA-----LTTPEPTPTTP-----KEPASTTPK 323
Db 415 QADPKAQEVPTPVKNPVKKYKPPWEVDDEDPVEEVKQPEAPAKTTPVLKRKEPAKDTA 474

QY 324 EP-----TP-TTIKSAPTPKAPATTTTKSAPTTTKAPATTTTKEPA----- 364
Db 475 KPATSKTPEPEKDKDPVKPRDSSPKVAAKPDQAQA-TPVKNPVKWRPPWEDDETPA 533
QY 365 -----PTTPKAPATTTTKEPAT-----TTKSAPATTPKPE-----APTTPKAPATTP 407
Db 534 DDVSKPTDAKKTPSLAKKDPAPAKESLKPADTKAPAKPRDPSPKKVPATPAEKKTTPVLA 593
QY 408 KE-----PAPTTPKPEPTTPKPEATTTKEPATTPK-EPAPTAPKKAP-----TTPKE 456
Db 594 KKEPAGPADSKTKEPEKSKPRDPSPKKAVPAKVPKTEVAPAAVKKPEPISKPKDTPAKK 653
QY 457 PAPTTPKPEPATTTTKEP-----SPTTPKPEPATTTTKSAPTTTKEPATTTTKEPATTP-KEP 511
Db 654 AEPNSPVVP-PTPVKNPVKWKPPWEDDDAPAKPVSLPEPEKK-TPVLAKKAPTKPDSEA 711
QY 512 SPTTTKEPATTPK--EPAPTTPKKAP--TTPKPEPATTTTKEPATTTTKEPATTPA-PKE 566
Db 712 AADPVSPSSKDPKLAKKAPVKPRDPSPMKAVPIKAPKT--EVPVAVVKPEPVAKSRD 769
QY 567 PAPTTPKETAPTTPKKLTPTTP-----EKLAPTTPKAPATTTPEELAPTTPPEP-----T 616
Db 770 PSPKKAAP--VVPPTPVKNPVKWKPPWEDDDAPAEVNVNPEPEKKTTPVLAKKT 826
QY 617 PTTPEEPATTPKAAAPNTPKPEPATTTTKEPATTP--PKEPATTTTKEPATTPKTATPTT 675
Db 827 PVKPRDPSKKAAPKAPSTKTDAVPVSVKKPEPVSKPEPKKAEPNPSVVP-----PTP 882
QY 676 LKEPATTTPKKAPKELAPT-TTKEPTSTSDKAPATTPKGTAPTTPKPEPATTTTKEPAT 734
Db 883 VKNPVKKW-KPWEDDDEPTEEVKKPSE--PEKKTTPVLAK-KEPEKPD-APKVAAPKPD 937
QY 735 TTPKGTAPTTLKEPAT-----TPKKAP-----KELAPTTPTKGPIS-----TTS DK 776
Db 938 PSPKKAAP--KEPAKVAAPKPDLSPKKAIPIPANTQEPATTPVKNPVKWKPPWEDDD 995
QY 777 P-----APTTPKET-----APTTPKPEPATTPKPKPATTPPTTPPTTSEVSTPTTKEP 825
Db 996 PAEPVSAPEPEKKTTPVLAKKAPAKPRDP---SPKKAAPVAAK-PDKPIEV-PPTPVKNP 1050
QY 826 TTIHKSP-----DESTPELSA-EP---TPKALENSKPEPGV-----PTTKTFA 864
Db 1051 VKWKPPWEDDDEPSEPVSAPEPEKKTTPVLAKKAPKATKPDSEAAADPVSGTSDPK 1110
QY 865 AT-----KPEMTTAKDKTTERDLRTTPETT--TAAPKMTKETATTEKTTESKITATT 916
Db 1111 LSKKAPVEKPKPTTDPKDDKLKPSAKKPEKAPKAPKWKPPWDDDDPDEPADEFTVPA 1170
QY 917 TQVTSSTTQDTPPKITTLTKTTTLAPKVTTTKTITTTTEIMNKPEETAKPKDRATNSKAT 976
Db 1171 PSKKPDTEPADPLG-----GPKTKDPK-----LNKKAPAEKPTK----- 1206
QY 977 TPKPKQKPTKAPKPTSTKPKTTPMRVKKPK-----TTPTRKMTSTMP 1019
Db 1207 -PKKEVSKPEPKPTPEPKP-AAPKWKPPWDDDDPDEPADEFTMPAPKPKDTEDP 1259

RESULT 5

T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C:Accession: T30826
R:Yotcov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A:Title: Differential splicing in of a proline-rich exon converts alphaNAC into a mus
A:Reference number: Z20889; MUID:96312450
A:Accession: T30826
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>

Mon Apr 29 08:35:16 2002

QY 61 SFCAEVHNTSPSPSKKAPPPGASOTIKSTTKRSPKPPNKKTKKVKVIESEBITEHSVS 120
Db 57 -----APSN-----PDVSTISS-----INNDNVDIGPSGD 83
QY 121 ENOESSSSSSSSSTIWKIKSSKNSAANRELQKLLKVKDNKKNKTKKPPKPPVWDE 180
Db 84 SNPTGSSFOEIEATEVGGQTVKSEHNIDSVEVEKV----- 120
QY 181 AGSLDNGDFKVTPTDTSTQ---HNKYSTSPKITTAKPINPRPSLPNPSDT----- 229
Db 121 -----TSTDASTNAPTCKDSTPPELLIGIIVINSKESVMDSTTFRESTTL 169
QY 230 SKETSIVNKETTVETKETTNTKOTSDCKEKTSAKETSQIEKTSADLAPTSLVLAK 289
Db 170 SPITELTSPETLVSDSSTST-EQTSPONTTEIASPMETNTTTEATTSVSPSVSTLAS 228
QY 290 PTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTTPKSAPTTPKEPAPTTKSPAP 349
Db 229 ---EDETTVTAIAEST-----TTVIAEVSTTTEEPITT-----AESTTKKST 267
QY 350 TTPKEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTTPKSAPTTPKEPAPTTTP-----KKPAPT 405
Db 268 T----KAPATTEPTPTTEE-VTTTEAETSTTSETSTEK---PTTPLIDNKKIAGPA 318
QY 406 TPKEPAPTTTPKEP---TPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTAPKAPPTTPKEPAPTTTPKE 464
Db 319 TGK---PETTHFPVTGTPN--FDATATETPFVAKSEDKWTLSTAAETTTQQTEVT--DG 372
QY 465 PAPTTEE---PSPTPKEPAPTTTPKSAPT-----TTTKPAPT-----TTKSAPTTP 508
Db 373 PEKETKKNVSEIETITVPLVETTSSTASKESDGFHTLLKLVTTADSDSTESATTVK 432
QY 509 KEPSPTTKE---PAPT-----TPK-----EPAPTTPKKPAP----- 537
Db 433 PNEETTTKSHVVPKPKTKGTVKVTPKLELSFDEPTEIT-KAPHPGKLLKKYHVFVLS 491
QY 538 ---TTPKE-----PAPTTPKEPAP-----TTPTKPAPTA--PKEPAPTT 571
Db 492 NFARYSEAKENDDYNHLDYNYHREAKEPTTTEESSTTEEVTTTEEPANTGNPPTTENPTT 551
QY 572 ---PKETAPTTPKKLTPTTPKELA---PTTPKEPAPTTPELATTPTEEPPTTPPEPAP 625
Db 552 TEQPTSTAESATTALPFTTEQTVTTEPTTAEKSTATO-----KPTTQESVST---BKTS 604
QY 626 TTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTPKETAPTTPKGTAPTTTLKEPAPTTPK 685
Db 605 TTKA---STTEE--PTTIDEPTTT---ESSTTGKATTPELSTTSEETTTTELKITTE- 655
QY 686 KPAPKELAPTTTKEPTSTSDKPAP-----TTPKGTAPTTTPKEPAPTTTPKEPAPTTPKGTA 741
Db 656 -----GSTTTEEPTTAIFAEASTGIITDEETSTSTTPELTSTKE---IVTESAIT 706
QY 742 PTLKEPAPTT-----KKPAPKELAPTTTKGP----- 769
Db 707 QTSVSVVESSTPRQLPERKAIIVNKFHNLVLEKELLKKEKSTSTSGSDSETTIV 766
QY 770 -----TSTSDKAPTTPKETAPTTTPKEPAPTTTPKAPPTPE--TTPPTTSE-VSTP 819
Db 767 AENIDEVTTTEKEKVQVTPITTEKSTTQETTTTTTTTTEKTSKTTTEKPTSESATTE 826
QY 820 TTTKEPTTHKSPDESTPELSAETPPKALENSKPEPGVPTTKTPAATKPEMTTAKDKTT 879
Db 827 TTTSEPT-----TEST-----TVDTSSATTEESSTAFTTTT 859
QY 880 EROLRTTTPETTA-----APKMKET 900
Db 860 SAE---TSETTTSAAFITGESPENTALQSSQKSEENESSAEKPGARRDFVPKKHKT 916
QY 901 ATTEKTESKITATTQO-VTST---TQDTTPFKITTLTKTTTLAPKV---TTTKTITT 953
Db 917 VKPAETTSAAVASTTTEPTTTEKSTTLETTTPIEATTLNEVTGPAFVGPVDETTINT 976
QY 954 TEIMNKPETAKPKDRATNSKATTPKPQ-----KPTKAPKKPTSTKKP 996

Db 977 LELLSK-----INNIQISQPKPDTISKDALSSSLIGSLGTSFKAPMAPTI----- 1022
QY 997 KTMPRVRRKPKTTPPRKWTSTMPELN-----PISRIAEA 1030
Db 1023 -----HTTDAAFVTATEASLNDGSKDKIIDEAQPTDEIRRA 1059
RESULT 10
A35175
mucin 1 precursor, repetitive splice form A [validated] - human
N:Alternate names: breast carcinoma-associated DF3 antigen; core protein KP39; episia
ncretic mucin; polymorphic epithelial mucin (PFM)
N:Contains: mucin 1 precursor, epithelial tumor antigen splice form; mucin 1 precursor
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 sequence, revision 20-Apr-2000 #text_change 02-Jun-2000
C:Accession: A35175; B35175; A35886; A35887; S10572; S40293; A36735; PX0066; S10218;
R:Litgenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilken, J.
J. Biol. Chem. 265, 5573-5578, 1990
A:Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene
A:Reference number: A35175; MUID:90202794
A:Accession: A35175
A:Molecule type: mRNA
A:Residues: 1-952,1033-1344 <LIG1>
A:Cross-references: GB:M32738; GB:J05288; NID:g182121; PIDN:AAA35804.1; PID:g182124;
A:Experimental source: splice form A
A:Note: GenBank entries HUMEPISIAL and HUMEPISIA2 present only the amino- and carboxyl
A:Accession: B35175
A:Molecule type: mRNA
A:Residues: 1-19,29-952,1033-1344 <LIG2>
A:Cross-references: GB:M32739; GB:J05288; NID:g182126; PIDN:AAA35806.1; PID:g182129;
A:Experimental source: splice form B
A:Note: GenBank entries HUMEPISIB1 and HUMEPISIB2 present only the amino- and carboxyl
R:Gendler, S.J.; Lancaster, C.A.; Taylor-Papadimitriou, J.; Duhig, T.; Peat, N.; Bure
J. Biol. Chem. 265, 15286-15293, 1990
A:Title: Molecular cloning and expression of human tumor-associated polymorphic epith
A:Reference number: A35886; MUID:90368715
A:Accession: A35886
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19,29-992,1033-1344 <GEN>
A:Cross-references: GB:J05581; NID:g188869; PIDN:AAA59876.1; PID:g188870
A:Note: GenBank entry HUMUCAB includes one copy of the tandemly repeated sequence
R:Lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.
J. Biol. Chem. 265, 15294-15299, 1990
A:Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.
A:Reference number: A35887; MUID:90368716
A:Accession: A35887
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19,29-1109,'S',1111-1339,'A',1341-1344 <LAN>
A:Cross-references: GB:J05582; NID:g189598; PIDN:AAAG0019.1; PID:g189599
A:Note: GenBank entry HUMPANMU contains four fewer copies of the tandemly repeated se
R:Wreschner, D.H.; Hareuveni, M.; Tsarfaty, I.; Smorodinsky, N.; Horev, J.; Zaretsky,
Eur. J. Biochem. 189, 403-473, 1990
A:Title: Human epithelial tumor antigen cDNA sequences. Differential splicing may gen
A:Reference number: S10571; MUID:90276413
A:Accession: S10572
A:Molecule type: mRNA
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1344 <WRE>
A:Cross-references: EMBL:X52229; NID:g37053
R:Wreschner, D.H.
submitted to the EMBL Data Library, March 1990
A:Reference number: S40293
A:Accession: S40293
A:Molecule type: mRNA
A:Residues: 1-19,29-155,'P',157-175,'P',177-182,'A',184-212,1033-1037,'A',1039-1344 <
A:Cross-references: EMBL:X52229; NID:g37053; PIDN:CAA36478.1; PID:g37054
R:Abu, M.; Siddiqui, J.; Kufe, D.
Biochem. Biophys. Res. Commun. 165, 644-649, 1989
A:Title: Sequence analysis of the 5' region of the human DF3 breast carcinoma-associat
A:Reference number: A36735; MUID:90088473
A:Accession: A36735

QY 1012 ---RKMSTMPENLNPSTGRIAEAMLQTTT 1036
Db 7806 KEIKDIIILTESEFVGSSAIFECIVSPST 7833
RESULT 12
T45025
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T45025
J:Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
R: Biol. Chem. 272, 3168-3178, 1997
A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter
A:Reference number: 222899; MUID:97166151
A:Accession: T45025
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3570 <DB>
A:Cross-references: EMBL:272496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A:Experimental source: placenta
C:Genetics:
A:Gene: MUC5B
Query Match 9.0%; Score 653; DB 2: Length 3570;
Best Local Similarity 22.1%; Pred. No. 7.3e-20;
Matches 394; Conservative 120; Mismatches 588; Indels 680; Gaps 71;
QY 22 VSSOELSCGRGCFEFGREGCDQAQCKYDKCCPDYE--SFCAEV-HNPTSPSSSKA 78
Db 1572 VPLRELGVVVEC--SLDFGLVCRNEOVGKF-KMCFNYEIRVFCNGYCHCPSTPATSTA 1628
QY 79 PPPS-----GASQIKSTTKRSPPNKKTKKVIIESEITEEHS 118
Db 1629 TPSPGCTTWILLTEQTAATATTATTGTAIPSTPGTAPPP-----KVLTQAATPTAT 1682
QY 119 VSENQSSS-----SSSSSSSSSTIWKIKS-----KNSAANR----- 151
Db 1683 SSRATSSSPRTATTPLVLTSTATKSTATSFPIPSSTLGTGTSONRPPHPMATMSTIH 1742
QY 152 -----ELQKLKYKDNKKNKTKKPKPP-----VVDE-----AGSLDNG----- 188
Db 1743 PSSTPETHSTVLTTKATTTRATSSMSTPSPGTTWILTTLTAATTAALPHGTPSS 1802
QY 189 -----DFKVTPTDSTTOHN-----KVSTSPKITAKPINR----- 220
Db 1803 TPGTTWILTEPSTATTVVTGSTATASSTRATAGTLKVLTS-TATPTVISSRATPSSS 1861
QY 221 -----PSLPNNSDTSKETSIVNKETTVET-----KETTTNKOTSTDGKEKTTSAKE 268
Db 1862 PGATAPALRSTATTPTATSVTAIPSSSLGTAWTRLQSQTTPATMSTATSPSTPETVH 1921
QY 269 TQSEIKTSADLAPTSKVLAKPTPKAETTK-----GALATP-----K 307
Db 1922 TSTVLTATTATRTGVSATPSPSTPGTAHTTKVPTTTTGTATPSSPGTALLPPVWIST 1981
QY 308 EPTPT-----TPKE-----PASTTPKEPTPTTIKSAPTT 336
Db 1982 TTPTTGTSTVTPSSIPGTTHTATVLTTLTTTATGSMATPSSSTQTSPTPSLTATTAT 2041
QY 337 PKEPATTTTSAKAP--TTPKEPATTTT--KEPAT--TTPKEPATTTTKEPATTTTSA-- 387
Db 2042 ITATGTTNPSTPGTTPPIPVLTTLTATTAASTSTVTPSSALGTTHTPPVNTTATHG 2101
QY 388 -----PTTP-----KEPATTPPKPATPT-----PKPATPTPKPAT 419
Db 2102 RSLPSSPHVPTAWTNTSATSGILGTHTEPSTCTSHTPAATTTGTQTPALSSPHSS 2161
QY 420 PTTKEPATTTKEPAT--TTPKEPATAPKAPATTKEPATTT--PKPATTTTKEPS-- 474
Db 2162 RTTESPPSPGTTTPGHTRGTSRTTATATPSPSKTRTSTLLPSSPTAPITVVVTTGCEFOCA 2221

QY 57 PYESFCAVHNPTSPSSKKAPPPSGASQTIKSTTKRS-----PKPPNKK 102
Db 6827 PEEEEVPE--EEVLPEVKVPVPAPVPEIKKKVTEKKWIPKKEEAPPKAKVPEVPKKV 6884
QY 103 KTKKVI--ESEEITEHVSVENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVK 160
Db 6885 EEKKIILPKEEEVLPVEVTEPEEIPSEE-----EIPPEPPSIEVEEVAPPRVP 6935
QY 161 DNKNKTKKKPTKPPVVDGAGSLDN--GDFKVTPP-DTSTQHNKVSSTPKIITAKP- 216
Db 6936 EVIKKAVPEAPTVPKKVCEAPPAKVKSKKIPBEKVVPVQKKKEAPKAVPEVKVPKVKV 6995
QY 217 -INPRPSLPP--NSDTSKETSIVNKETTVFK-----ETTTNNKQTSDDGKE----- 261
Db 6996 LVPKKEAVPPAKGRTVLEEKVSVAEQEVVVKERLEVEVEAEVIEPEEEFHEVEEYF 7055
QY 262 -----KTSAKETQSIEKT-----SAKDLAPTSSKVLAKPTPKAETTTKG 300
Db 7056 EEEGFHEVEEFIKLEQHRVEEHRVEKVKHVIEVFAGEVEVEKPKAP-----KG 7107
QY 301 PALT-----TPKEPTPTTP--KEPASTTTPKEPTPTTIKSAPTTTPKEP-APTTKSAPTTPK 353
Db 7108 PEISEKLIIPKPKPTKVVRKPEKPAKVPVPPKKIVVEKVRVDEPRVPPTKVPEVLPK 7167
QY 354 EPARTTTKEPATTPKPEPATTTTKEPATTTTKSAPTTPKEPATTPPK-----PAPTTPKE 409
Db 7168 EVVP-----EKKVVPVPAKKPEAP-----PPKVPEAPKEVVPKVPVPPKKPEV 7213
QY 410 PAPTTPKEPTTPKPEPATTTKPEPATTP-----KEPATAPKPAPTTPKPEPATTP 462
Db 7214 PPTKVPEVPAKAAVDEKVKVPEAIPKPKESPPEVEEPEESPSAPPK-KPEVPPVRVPEV 7272
QY 463 KEPATTTTKEPSPTTPKEPATTTTKSAPTTPKEPATTTTKSAPTTPKEPSPTTPKEPAT 522
Db 7273 KEVVPKVKPAAP--PKKPEVTPVK-VPEAPKEVVPKVKVPVPP--PPKKPEVPTK----- 7323
QY 523 TPKEPATTPPKKPAPTTPKPEPATTPPKKPAPTTPKAPKAPKAPTPTPKETAPTTPPK 582
Db 7324 VPEVKVAVPEKKVP-----EALPKPESPPEVEEPEEVALEBPAAVEVEPEEPAAPPQ 7379
QY 583 LT-----PTTPKPLAPT-----PEKAPTTPELAPTTPPEEPTP-TTPEEPAPTTPKAAAP 633
Db 7380 VTVPPKNDVPEKKAPAVVAKKPELPELVKVPVPEVPEKVPVPEKKVPLVVPKK--PEAPPKVP 7437
QY 634 NTPKEPAP--TTPKEP-----APTTPKEPATTPKETAAPTTPKGTAPTTLKEPATPT 683
Db 7438 EYPKPEVPEKKVAVPKKPEVPPAKVPEVPKPV-----LEEKPAVP--VPERAESPPPEV 7490
QY 684 PKKPAKELAPTTPKPTSTTSKAPPTPKGTAPTTPKPEPATTPKPEPATTPKGTAPT 743
Db 7491 YEE--PEBIAP---EEELAPEEEKVPVVAEE-----EEPEVPPAVPEEPKKIIEKKVP- 7540
QY 744 TLKEPATTPKPAKELAPTTPKGTPTSTSDKPAPTTPKETA-----PTTPKEPATPT 796
Db 7541 VIKKPEAPPKPEPEKVP-----IEKPKLPRPPPPPPAPPKEDVKELIQLKAIKPKVPE 7597
QY 797 TPKKP--APTTPETPPPTSEVS--TPTTPKEP-----TTIHKSPDESTPELSAETP 845
Db 7598 NQOVPEKVELTFLKVPGGKKVKRLPERKPEKPEEVVLKSVLRKRPEEPEEFPKEV---P 7653
QY 846 KALENSKPEGVPPTTKTPAATKP-----EMTTAKDKTITERDLRTTPETTTAAPKMTKETA 901
Db 7654 KKLE-KVKKPAVP---EPPPKPVEVEVPTVTKRKELIPEPTKVPEIKPAIPLPAPBPK 7709
QY 902 TTTEKTTESKITATTOVTTSTTQDTTPFKIITLTKITTLAPKVTTTKTITTTTEIMNK-- 959
Db 7710 PKPEA-----EVTIKPPVPEPPTPIAAPVTVPVVGGKAAE 7745
QY 960 -----PEETAKPKDRATNSKATTPKQKPTKAPKPKTS--TKPKTMVRVRKPKITPTP-- 1011
Db 7746 AKAPKEAAKPKGPIKGVPKKTPSPJEAERKKLRPFSGGKEKPPDEBAPFTYQLKAVPLKPV 7805

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:32:03 ; Search time 62.75 Seconds
(without alignments)
796.401 Million cell updates/sec

Title: AA2
Perfect score: 7276
Sequence: 1 MAWKLPDYLLLSVFIQ.....ARAITTRSGOTLSKVMWVNC 1363

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1175	16.1	5179	1 MUC2_HUMAN	Q02817 homo sapien
2	950	13.1	1664	1 SLPI_CLOTM	O06852 clostridium
3	809	11.1	1367	1 AMYH_YEAST	P08640 saccharomyc
4	651	8.9	1255	1 MUC1_HUMAN	P15941 h mucin 1 p
5	559	7.7	875	1 FPL_MYRD	Q25460 mytilus edu
6	555.5	7.6	2700	1 ZAN_HUMAN	Q9Y493 homo sapien
7	551	7.6	620	1 EXTN_TBOAC	P13983 nicotiana t
8	551	7.6	1087	1 NFM_MOUSE	P19246 mus musculu
9	530.5	7.3	1162	1 TCNA_TRYCR	P23253 trypanosoma
10	530	7.3	865	1 CPN_DROME	O02910 drosophila
11	518.5	7.1	872	1 FPL_MYCO	Q25343 mytilus cor
12	517.5	7.1	831	1 NFM_MOUSE	P16884 rattus norv
13	509.5	7.0	662	1 MUC1_XENLA	Q05049 xenopus lae
14	503.5	6.9	1970	1 RPB1_HUMAN	P24928 homo sapien
15	502	6.9	1970	1 NFM_MOUSE	P13983 nicotiana t
16	498.5	6.9	1970	1 RPB1_MOUSE	P08775 mus musculu
17	493.5	6.8	467	1 RPB1_CRIGR	P14144 cricetus
18	490	6.7	2142	1 BART_HUMAN	P46834 homo sapien
19	488.5	6.7	826	1 SRR2_PLAYO	Q01443 plasmodium
20	475.5	6.5	267	1 EXTN_MAIZE	P14918 zea mays (m
21	471.5	6.5	5376	1 ZAN_MOUSE	O08799 mus musculu
22	468.5	6.4	634	1 HWP1_CANAL	P46593 candida alb
23	467	6.4	817	1 VRL1_YEAST	P37370 saccharomyc
24	454	6.2	797	1 VGLX_HSEVB	P28968 equine herp
25	454	6.2	1161	1 YJ9P_YEAST	P47179 saccharomyc
26	453.5	6.2	1794	1 VGLX_HSEVB	Q010130 schistosach
27	448.5	6.2	670	1 VGLX_HSEVB	Q010130 schistosach
28	446	6.1	1229	1 VGLX_HSEVB	Q010130 schistosach
29	444.5	6.1	1229	1 VGLX_HSEVB	Q010130 schistosach
30	442.5	6.1	1229	1 VGLX_HSEVB	Q010130 schistosach
31	439.5	6.0	1083	1 FPL_MYGA	P34926 rattus norv
32	436.5	6.0	1411	1 TCDF_HUMAN	O27409 mytilus gal
33	434.5	6.0	3256	1 KIT67_HUMAN	O13428 homo sapien

34	433.5	6.0	3164	1 TEGU_HSV11	P10220 herpes simp
35	432.5	5.9	439	1 XP2_XENLA	P17437 xenopus lae
36	428	5.9	2517	1 NCR2_HUMAN	Q9Y618 h nuclear r
37	427.5	5.9	1251	1 YOUN3_CAEEL	O09550 caenorhabdi
38	426	5.9	2715	1 TRX2_HUMAN	O9umh6 homo sapien
39	425.5	5.8	1185	1 DRPL_HUMAN	P25459 homo sapien
40	424.5	5.8	1125	1 MAP4_MOUSE	P27546 mus musculu
41	424	5.8	307	1 SG53_DROME	P02840 drosophila
42	423.5	5.8	3421	1 TEGU_HSEVB	P28955 equine herp
43	422.5	5.8	1183	1 DRPL_RAT	P54258 rattus norv
44	419.5	5.8	2476	1 ZAN_PIG	O28983 sus scrofa
45	411.5	5.7	907	1 VGP3_EBV	P03200 Epstein-bar

ALIGNMENTS

RESULT 1
MUC2_HUMAN STANDARD: PRT: 5179 AA.
AC Q02817: Q14878:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
CN MUC2 OR SMOC.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE-94132002; PubMed-8300571;
RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor."
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE-93016075; PubMed-1400449;
RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region."
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RC MEDLINE-91358717; PubMed-1885763;
RA Toribara N.W., Gum J.R., Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism."
RL J. Clin. Invest. 88:1005-1013(1991).
CC -1- FUNCTION: CODES THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIONS
CC AGENTS AT MUCOSAL SURFACES.
CC -1- SUBUNIT: MULTIMERIC.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -1- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
CC VARIES AMONG DIFFERENT ALLELES.
CC -1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC OF SLIKORM HEMOCYTIN.
CC -1- SIMILARITY: CONTAINS 2 WVEC DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTEINE KNOT-LIKE DOMAIN (CTCK).
CC

Query Match	Best Local Similarity	Score	DB 1:	Length	DB 2:	Score	DB 3:	Length	DB 4:	Score	DB 5:	Length	DB 6:	Score	DB 7:	Length	DB 8:	Score	DB 9:	Length	DB 10:	Score	DB 11:	Length	DB 12:	Score	DB 13:	Length	DB 14:	Score	DB 15:	Length	DB 16:	Score	DB 17:	Length	DB 18:	Score	DB 19:	Length	DB 20:	Score	DB 21:	Length	DB 22:	Score	DB 23:	Length	DB 24:	Score	DB 25:	Length	DB 26:	Score	DB 27:	Length	DB 28:	Score	DB 29:	Length	DB 30:	Score	DB 31:	Length	DB 32:	Score	DB 33:	Length	DB 34:	Score	DB 35:	Length	DB 36:	Score	DB 37:	Length	DB 38:	Score	DB 39:	Length	DB 40:	Score	DB 41:	Length	DB 42:	Score	DB 43:	Length	DB 44:	Score	DB 45:	Length	DB 46:	Score	DB 47:	Length	DB 48:	Score	DB 49:	Length	DB 50:	Score	DB 51:	Length	DB 52:	Score	DB 53:	Length	DB 54:	Score	DB 55:	Length	DB 56:	Score	DB 57:	Length	DB 58:	Score	DB 59:	Length	DB 60:	Score	DB 61:	Length	DB 62:	Score	DB 63:	Length	DB 64:	Score	DB 65:	Length	DB 66:	Score	DB 67:	Length	DB 68:	Score	DB 69:	Length	DB 70:	Score	DB 71:	Length	DB 72:	Score	DB 73:	Length	DB 74:	Score	DB 75:	Length	DB 76:	Score	DB 77:	Length	DB 78:	Score	DB 79:	Length	DB 80:	Score	DB 81:	Length	DB 82:	Score	DB 83:	Length	DB 84:	Score	DB 85:	Length	DB 86:	Score	DB 87:	Length	DB 88:	Score	DB 89:	Length	DB 90:	Score	DB 91:	Length	DB 92:	Score	DB 93:	Length	DB 94:	Score	DB 95:	Length	DB 96:	Score	DB 97:	Length	DB 98:	Score	DB 99:	Length	DB 100:	Score	DB 101:	Length	DB 102:	Score	DB 103:	Length	DB 104:	Score	DB 105:	Length	DB 106:	Score	DB 107:	Length	DB 108:	Score	DB 109:	Length	DB 110:	Score	DB 111:	Length	DB 112:	Score	DB 113:	Length	DB 114:	Score	DB 115:	Length	DB 116:	Score	DB 117:	Length	DB 118:	Score	DB 119:	Length	DB 120:	Score	DB 121:	Length	DB 122:	Score	DB 123:	Length	DB 124:	Score	DB 125:	Length	DB 126:	Score	DB 127:	Length	DB 128:	Score	DB 129:	Length	DB 130:	Score	DB 131:	Length	DB 132:	Score	DB 133:	Length	DB 134:	Score	DB 135:	Length	DB 136:	Score	DB 137:	Length	DB 138:	Score	DB 139:	Length	DB 140:	Score	DB 141:	Length	DB 142:	Score	DB 143:	Length	DB 144:	Score	DB 145:	Length	DB 146:	Score	DB 147:	Length	DB 148:	Score	DB 149:	Length	DB 150:	Score	DB 151:	Length	DB 152:	Score	DB 153:	Length	DB 154:	Score	DB 155:	Length	DB 156:	Score	DB 157:	Length	DB 158:	Score	DB 159:	Length	DB 160:	Score	DB 161:	Length	DB 162:	Score	DB 163:	Length	DB 164:	Score	DB 165:	Length	DB 166:	Score	DB 167:	Length	DB 168:	Score	DB 169:	Length	DB 170:	Score	DB 171:	Length	DB 172:	Score	DB 173:	Length	DB 174:	Score	DB 175:	Length	DB 176:	Score	DB 177:	Length	DB 178:	Score	DB 179:	Length	DB 180:	Score	DB 181:	Length	DB 182:	Score	DB 183:	Length	DB 184:	Score	DB 185:	Length	DB 186:	Score	DB 187:	Length	DB 188:	Score	DB 189:	Length	DB 190:	Score	DB 191:	Length	DB 192:	Score	DB 193:	Length	DB 194:	Score	DB 195:	Length	DB 196:	Score	DB 197:	Length	DB 198:	Score	DB 199:	Length	DB 200:	Score	DB 201:	Length	DB 202:	Score	DB 203:	Length	DB 204:	Score	DB 205:	Length	DB 206:	Score	DB 207:	Length	DB 208:	Score	DB 209:	Length	DB 210:	Score	DB 211:	Length	DB 212:	Score	DB 213:	Length	DB 214:	Score	DB 215:	Length	DB 216:	Score	DB 217:	Length	DB 218:	Score	DB 219:	Length	DB 220:	Score	DB 221:	Length	DB 222:	Score	DB 223:	Length	DB 224:	Score	DB 225:	Length	DB 226:	Score	DB 227:	Length	DB 228:	Score	DB 229:	Length	DB 230:	Score	DB 231:	Length	DB 232:
-------------	-----------------------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------	-------	---------	--------	---------


```

OY 793 PAPTPKPPAPT--TPPTPTTSEVSTPTTKETTHKSPDESPBELSAEPTRKALENS 851
DB 1254 PTPS--DEPPSPETPEPTITDPSDEPTSDPT-----PSDEPT--SDEPTPSDEPT 1305
OY 852 PKRGVPTTKPPAT--KPEMTTAKDKTERDITPTETTTAAPKMTKATTAITEKTE 909
DB 1306 SDEPTPSDEPTSEPTPEPTITDPSDEPTSD-----EPSPDEPTPSDEPTSDPTPS 1361
OY 910 SKITATTTQVSTTTQDPTTFKTTTLTKTLAPKVTTKTKTTTTEIMNKPEETAKPKDR 969
DB 1362 DEPTSPDEPTPEPTPEPT-----TTPTPTPTT-----PTSG 1396
OY 970 ATNSKAT-----TPKPKPTKAP--KKPTSKPKPTMRVRKPTTPPKM----- 1014
DB 1397 SGSGSGSGGGGGGGGVPSPPTPTPSKPTSPAP--TEIEPTPSDVPCAIGCEHRA 1453
OY 1015 -----TSMPELNPTSRIAEAMLOTTTRNQTTPNKNKLVENKPSDAGAGEPTPHM- 1066
DB 1454 YIRGYPDSFSPERKNTIRAEAAVIF-----AKLL-----GADESTGAQASAFYSD 1498
OY 1067 LLRP-----VEMPE-----VTPMDYLPVYVNOGI----- 1092
DB 1499 LAOTHHMAAIAKFAKSGLEFGYDGTFFKPPONTIRAEFAVVLHFLTKVKGCEIMSKLA 1558
OY 1093 ---IINPMLSDETNINCG---KPYDGLTV-----RNGTLVAFRGHYFMMLSPSPSPS 1139
DB 1559 TIDISNPFKFD---CVGHMAOEFEKLTLSGYISGYPDGT-----FKPON 1600
OY 1140 PARRTTEWAGIPSPIDVFTRCNCGKTFPFKD--SOYWRFTNDIKD 1184
DB 1601 YIKRESV-----ALINRALERGLNGAKPLFPDVNESYWA- GDIMD 1642

RESULT 3
AMYH_YEAST STANDARD: PRT: 1367 AA.
AC P08640: P08068 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUCOAMYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-
DE GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).
CN STA1 OR STA2 OR MAL5 OR YIR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
OY 11
RN
RN SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Geniesse S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Raebig M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RA Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN 12
RN SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RP MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STA1";
RL J. Bacteriol. 169:2142-2149(1987).
RN 13
RN SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar elements in the 5' regions of the STA2 and SGA genes
RT from Saccharomyces cerevisiae";
RL FEBS Lett. 239:179-184(1988).

```

```

CC -! CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
CC GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
CC WITH RELEASE OF BETA-D-GLUCOSE.
CC -! SIMILARITY: TO S.POMBE SPC215.13.
CC -! SIMILARITY: SOME, TO S.POMBE SPC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: Z38061; CAAB6176.1; -
DR EMBL: M16164; AAA35014.1; -
DR EMBL: M16165; AAA35015.1; -
DR EMBL: X13857; CA332069.1; -
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR PIR: S48478; S48478.
DR SGD: S0001458; MOC1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT CHAIN 1 21
FT DOMAIN 22 1367
FT CARBOHYD 817 1367 SER/THR-RICH.
FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1367 AA: 136110 MW: 9100E2BDBD5LIA9D CRC64;

```

Query Match 11.1%; Score 809; DB 1; Length 1367;

Best Local Similarity 27.5%; Pred. No. 9,8e-25; Matches 328; Conservative 112; Mismatches 526; Indels 226; Gaps 53;

```

OY 44 DCDAAQKKYK---CCPDYSCFAEYHNPTSPSSSKAPPPSCASQITSTKRSKPPN 100
DB 174 DLSTGCNNYDNOGHSGTDEPGFYWNIDCNGCGTSSITTSSTSS--ST----- 223
OY 101 KKKTKVIESEETIEHVSSENOESSSSSSSTIMKIKSKNSAANRELQKLKYK 160
DB 224 ---TSSISESSTTSSSESSTTSSSESSTST-----APATPTVTSCTKKEKPP 274
OY 161 DNKKRTRKKRPPKPPVVDAGSLDNGDFKVTTPOTSTTQHMKVSGSPKTTAKPINR 220
DB 275 PTTTSCYKREKPT--PRHND-----TTPCT-----KKTTTSTCKTKKTTTTPY 314
OY 221 PSLPPNSOTSKESTLVNKEETVETKETTNTKQSTDKETTSKAKETOSIKTSKADL 280
DB 315 PT--PSSSTTSSS-----APVTPSSSTT-----ESSAPVTSSTTSSS--- 353
OY 281 APTSKVLAKPPPKAETTKGPAITTPKEPPTTPKEPASTTPKEPPTTKSAPTTPKEP 340
DB 354 APV-----PTPSSSTTSSSAPVTSSTTSSSAPVTSSTTSSSAPVPTPSSSTTSSS 407
OY 341 APTT-----TSSAP-----TTKEPAPTTPKEPAPTTPKEPAP-----TTKEPAPTTPKSA 387
DB 408 APVTSSTTSSSAPVTSSTTSSSAPVTS-----TSSSAPVTSSTTSSSAPVTPPSS 463
OY 368 PTPPEPAPTTPKPPAPTTPKEPAPTTPKEPTP---TPKEPAPTTPKEPAPTTPKEPAPT 444
DB 464 STTSSSAPV-----STTSSSAPV-----PTPSSSTTSSSAPV-----SSTTSSSAPV 512
OY 445 APKKPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSA 504
DB 513 -----PTPSSSTTSSSAPVTPS---SSTTSSSAPVTS---TSSSAPVTPPSS 559
OY 505 PTPPEPAPTTPKEPAPTTPKEPAPT--TPKKPAPTTPKEPAPTTPKPPAPTTPKPPAPT 563
DB 560 STTSSSAPVTPS-----TSSSAPVTPPSSSTTSSSAPVPT-----PSSSTTSSSAPA 611
OY 564 PPEPAPTTPKETAAPTTPPKLPTTPPEKLAAPTTPKEPAPTTPPELAAPT--TPEPTPTTPEE 622

```

```

Db      612 PRESSSTESSAPVT-----SSTESSAP-VPTPSSSTESSAPVPTPSSSTESSA 666
Qy      623 PAP-----TTPKAAANTPEKAPATTPKAPATTPKAPATTPKATPTL-TPKATPTTL 676
Db      667 PVPPTSSSTESSAPVT-----SSTESSAPVT-----SSTESSAPVPTPSSSTESS 718
Qy      677 KEAPATTPKAPKELAPATTPKATPTSTSDKAPATTPKATPTKAPATTPKAP 733
Db      719 SAPVPTSSSTESSAPVPTPS-SSTESSAPVPTSSSTESSAPVPTPSSSTESSA 777
Qy      734 PT-TKGTGAPTLKBPATTPKAPKELAPATTPKATPTSTSDKAPATTPKATPTPE 792
Db      778 PVPPTSSSTESSAPVPTPSSSTESSAPVPTPSSSNTSSAPSTPSSSTESSV 837
Qy      793 PAP-----TTPKAPATTPETPTSEKPT-----TKKPTTHKSPDESTPELSAE 842
Db      838 PVPPTSSSTESSAPVPTSSSTESSAPVPTPSSSNTSSAPSTPSSSTESS 894
Qy      843 PTPKALENSPKRPVPTTKTPATKPEMTTAKDRTTPTTETTAAPKMTKETAT 902
Db      895 GT-TVTPSSSKVPSQTEFSVSTETTVPTKTTTSVTPTSTTTTTCSTGTSNAGE 953
Qy      903 TTEKTESKITAT--TTOVSTTODTTPFKITLTKTTTLPKVT--TKITTTT-EI 956
Db      954 TTSKGSPTVTTVPTTTTSTTTTSTTTTTCSTGTSNAGESTGCSKPTTTTTPC 1013
Qy      957 MKKPEETAKPKDRATNSKATTPKPKP-----TKAPKPKSTKPKPKPTM 999
Db      1014 STSPET-----SESTTSTPTPTVSTTVSTTVTSTSTKGGELTTTFVKNI 1065
Qy      1000 PRVRPKPTTTPRKMTSTMPLNPTSRFAEMLOTTRPNOTPNKSLVENPKSEDA 1059
Db      1066 PTTTLTITAPTP--SVTTVTFPTT-TTTCVSTGT-----NSAGETTS 1109
Qy      1060 EGEPHMLLRPHVMEVTPDMDYLPRVNOGIIINPMLSDETNICNGKPVDTGLTRNG 1119
Db      1110 SPKVTVTTPCSTGTGTEATLTAVTAVTTVTTESSTGNNSA-GKTTGTTKTSVP 1168
Qy      1120 TLVAFRGHYEMLSFSPSPARITVWGIKPSPIDVTFRC-----NCEBKT 1167
Db      1169 TT-----YTTTLAPSAVTPATN-----AVPTTTT--TECSAATMAAGET 1207

```

```

RESULT 4
MUC1_HUMAN STANDARD: PRT: 1255 AA.
AC P15941: P15942; P13931; P17626; Q14128; Q16442; Q16437; Q9Y4U2;
DT 01-JUN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)
DE (EPITHELIAL) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)
DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG) (PEANUT-
DE REACTIVE URINARY MUCIN) (PUM) (BREAST CARCINOMA-ASSOCIATED ANTIGEN
DE D33).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=90368716; PubMed=2394722;
RA Ian M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90202794; PubMed=2318825;
RA Ligenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Episialin, a carcinoma-associated mucin, is generated by a

```

```

RT polymorphic gene encoding splice variants with alternative amino
RT termini.";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Peat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of human tumor-associated
RT polymorphic epithelial mucin";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Peat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed vntR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;
RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N.,
RA Horov J., Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A.,
RA Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horov J.,
RA Zilhan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transcribed gene and over-expression in breast cancer
RT tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]
RP SEQUENCE FROM N.A.
RX MEDLINE=91033045; PubMed=1688329;
RA Tsarfaty I., Hareuveni M., Horov J., Zaretsky J., Weiss M.,
RA Jeltsch J.M., Gendler J.M., Lathe R., Keydar I., Wreschner D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RL Gene 93:313-318(1990).
RN [8]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88330762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [9]
RP SEQUENCE OF 1-169 FROM N.A.
RX MEDLINE=90084473; PubMed=2597151;
RA Abe M., Stidiqui J., Kufe D.;
RT "Sequence analysis of the 5' region of the human DF3 breast
RT carcinoma-associated antigen gene.";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [10]
RP SEQUENCE OF 1-109 FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene.";
RL Int. J. Cancer 66:55-59(1996).
RN [11]

```

FT	CARBOHYD	1133	1133	N-LINKED (GLCNAC.) (POTENTIAL).
FT	VARSPLIC	19	19	T-> TATTAPKAPT (IN ISOFORM B).
FT	VARSPLIC	20	22	MISSING (IN ISOFORM C).
FT	VARSPLIC	20	31	MISSING (IN ISOFORM D).
FT	VARSPLIC	126	905	MISSING (IN POLYMORPHIC EPITHELIAL ISOFORM).
FT	VARSPLIC	1077	1087	FLOYKGGGFL -> VSIGISFPMPLP (IN SECRETED ISOFORM).
FT	VARSPLIC	1088	1255	MISSING (IN SECRETED ISOFORM).
FT	CONFLICT	2	2	T-> A (IN REF. 11).
FT	CONFLICT	134	134	P -> Q (IN REF. 9).
FT	CONFLICT	154	154	P -> Q (IN REF. 9).
FT	CONFLICT	1021	1021	S -> T (IN REF. 3).
FT	CONFLICT	1251	1251	A -> T (IN REF. 3).
SO	SEQUENCE	1255 AA:	122072 MM:	5E28DFC4DE7D9A82 CRC64.
Query Match				
Best Local Similarity 27.8%; Pred. No. 1e-18; Length 1255;				
Matches 295; Conservative 72; Mismatches 460; Indels 234; Gaps 51				
OY	23	TSLTVNKETVERKEETTTNKKSTGDKKTSKAKETOSIKTSKADP-LAPTSKVLAKPT	291	
Db	16	TVLTV-----VTGSGHASTGTGGEKKEISAAQSRSSVSSTKEKNASVTSVLSHSS	65	
OY	292	P-KAEVTTTKGP--ALTTPKRP-----TPTTKPEKASTPKKEPTPTTI--K	331	
Db	66	PGGSGSTTGGODVTLATETEPASGSAATWGODVTSVPTLRPALGSTTPPAHDVTSAPDNK	125	
OY	332	SAPTTKPEKAPATTTKKSAPTTKPEKAPATTTKPE-----APTTKPEAPATTTKPEAPTTTK	385	
Db	126	PAPGSTAPRAHGVTSAPDT--RPAQGSTAPRAHGVTSAPDT--RPAQGSTAPRAHGVTSAPDT--	179	
OY	386	SAPTTKPEKAPATTP-----KKRAPTTKEP-----APTTKKEPTPTT-----	423	
Db	180	SAPDTRRAPAGSTAPRAHGVTSAPDTRRAPAGSTAPRAHGVTSAPDTRRAPAGSTAPRAHGVTS	239	
OY	424	-----KEPAPTTKEP-----APTTKEPAPTTK-----KRAPTTKPEAPTTK-----E	464	
Db	240	SAPDTRRAPAGSTAPRAHGVTSAPDTRRAPAGSTAPRAHGVTSAPDTRRAPAGSTAPRAHGVTS	299	
OY	465	PAPTTKPEKAPTTTKPEKAPTTT-----TKSAPTTTKEAPTTTKSAPTTTKPEKAPTTTKPE	519	
Db	300	SAPDTRRAPAGSTAPRAHGVTSAPDTRRAPAGSTAPRAHGVTSAPDT--RPAQGSTAPRAH	356	
OY	520	-----APTTKPEKAPTTK-----KRAPTTKPEAPTTTK-----EPAPTTTKKRAPTTARKEP	567	
Db	357	GVTASAPDTRRAPAGSTAPRAHGVTSAPDTRRAPAGSTAPRAHGVTSAPDTRRAPAGSTAPRAH	416	
OY	568	APTTKPEKAPTTTKKLTPTTPEKLAPTTKEKRAPTTTPELAPTTTPEEPTTPEEAPTT	627	
Db	417	GVTASAPDTRRAPAGSTAPRAHGVTSAPDTRRAPAGSTAP--RAHGVTSAPDTRRAPAGSTA	472	
OY	628	PKA-----AAPPTKPEAPTTTK-----EPAPTTKPEKAPTTTKPEKAPTTTKGAPTTTLKEP	679	
Db	473	PRAHGVTSAPDTRRAPAGSTAPRAHGVTSAPDTRRAPAGSTAP--RAHGVTSAPDTRRAPAR	528	
OY	680	APTTP-----KKAPKELAPTTTKKEPTSTISKAP--TPKKGATPTTKEEAPTT	727	
Db	529	GSTAPRAHGVTSAPDTRRAPAGSTAP--RAHGVTSAPDTRRAPAGSTAPRAHGVTS-----APD	583	
OY	728	TPKEAPPTTKG-----TAPTTKEKAPTTT-----KKAPKELAPTTTKGPTSTT	773	
Db	584	TPRAPGSTAPRAHGVTSAPDTRRAPAGSTAPRAHGVTSAPDTRRAPAGSTAP--RAHGVTSAP	642	
OY	774	SDKPAP--TTPK-----ETAPTTKPEKAPTTP-----KKRAPTTPEPTTPTTSEVS	817	
Db	643	DTRAPAGSTAPRAHGVTSAPDTRRAPAGSTAPRAHGVTSAPDTRRAPGS--TAPRAHGVTS	700	
OY	818	TTTTTKEPT-----IKSPDESTPLSLSEPTPKALESSPKPEGVPTTKTPAATKPEM	870	
Db	701	APDTRRAPAGSTAPRAHGVTSAPDTRRAPAGSTAPRAHGVTSAP-----DTRAPAGSTAP--	753	


```

QY 871 TTAAADKTTENDLRTTPEPTT-----TAAPKMKETATTEKTESKITATTOVSTST 923
DB 754 --PAGVTSAPDTPRAPGTSAPRAHGVTSAP-----DTAPAGGASTAPRAHGVTSAPDTRPA 807
QY 924 TODTP--FKITTLTKTTLAKVTTTKTITTT--EINNKKEEPAKPKDRAATNSKATTPK 979
DB 808 PGSTAPRAHGVTSAPDTPRAPGTSAPRAHGVTSAPDTRPAGGASTAPRAHGVTSAPDTRPA 867
QY 980 POKPTKARKKPTSTKPKPTMPVRKPKPTPTPKMTSTMPKPLNTSRIAEMLQTTTRPN 1039
DB 868 P--GSTAPRAHGVTSAPDTRP--APGSTAPRAHGVTSAPDTRPA-----PG 909
QY 1040 QT-PRSKLVEVNPKESEDAGAEGETPHMLLRPHVMEVTPEDMDYLRVNPQGITINPML 1098
DB 910 STAPRAHGVTSAPDTPRAPGTSAPRAH-----GVTSAPDNRPALGSTA-----PPV 955
QY 1099 SDETNICNGKRPVDTLTTLRNGTLVAFNGHYFWMLSPEFPS 1139
DB 956 HNVTSASGSASGASTLVHNGTSARATTTTAPSKSTPSIPS 996

RESULT 5
FPL_MYTED STANDARD; PRT: 875 AA.
AC Q25460;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADHESIVE PLAQUE MATRIX PROTEIN (POLYPHENOLIC ADHESIVE PROTEIN) (FOOT
  DE PROTEIN 1) (MEP1) (FRAGMENT).
GN FPL.
OS Mytilus edulis (Blue mussel).
OC Eukaryotes; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
  OC Mytiloidea; Mytilidae; Mytilus.
  NCBI_taxid=6550;
  RN [1]
  RP SEQUENCE FROM N.A.
  RP MEDLINE=91025829; PubMed=1367451;
  RA Filippa D.R., Lee S.M., Link R.P., Strausberg S.L., Strausberg R.L.;
  RT "Structural and functional repetition in a marine mussel adhesive
  RT protein."
  RN Biotechnol. Prog. 6:171-177(1990).
  RL [2]
  RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
  RP MEDLINE=83135732; PubMed=6298211;
  RA Waite J.H.;
  RT "Evidence for a repeating 3,4-dihydroxyphenylalanine- and
  RT hydroxyproline-containing decapeptide in the adhesive protein of the
  RT mussel, Mytilus edulis L."
  RL J. Biol. Chem. 258:2911-2915(1983).
  CC -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
  CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
  CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
  CC -1- SUBCELLULAR LOCATION: SECRETED.
  CC -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
  CC -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
  CC -1- PTM: THE DECAPEPTIDE A-K-P-S-Y-P-P-T-Y-K IS POST-TRANSLATIONALLY
  CC MODIFIED AS FOLLOWING: THE SIXTH AND SEVENTH RESIDUES ARE
  CC HYDROXYLATED AND THE PENULTIMATE IS A 3,4-DIHYDROXYPHENYLALANINE
  CC (DOPA) DERIVED FROM TYROSINE.
  CC -----
  CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@sib-sib.ch).
  CC -----
  DR EMBL: X54422; CA838294.1; -
  DR InterPro: IPR002964; Adhesive_plaq.
  DR InterPro: IPR002965; P_flich_extensn.

```

```

DR PRINTS; PRO1216; ADHESIVEI.
DR PRINTS; PRO1217; PRICHEXTENSN.
KW Repeat; Hydroxylation.
FT NON_TER 1
FT DOMAIN 67 870
SQ SEQUENCE 875 AA: 100412 MW: 65E85312748CAGE CRC64:
P-P-(ST)
TANDEM REPEATS OF Y-K-[PA]-K-[LP]-(ST)-Y-
P-P-(ST)
Query Match 7.7%; Score 559; DB 1; Length 875;
Best Local Similarity 26.9%; Pred No. 2.6e-15;
Matches 281; Conservative 102; Mismatches 403; Indels 260; Gaps 56;
QY 67 HNPSTPSSSKKAPPPSGASQTIKSTKRSPPKPKKTKKVESEETIEHSVSENOESS 126
DB 3 HEVYKP-----VTSYSAKYKPPYQPLKKVD-----YRPT 35
QY 127 SSSSSSSSTTIWKIKSSNKAANRELQKKLVKONKKRTRKKLPKPPVVDGSGGLD 186
DB 36 KSYPTTYSKTNLPLAKLSS-----YKPIKTYNAK-----TNVPPVY----- 75
QY 187 NGDFKVTPTDSTTOHNVKSTSPKITTAAPINRPSLPNSDTSKETSILYKKEETVENK 246
DB 76 -----KPKMTYPTTKKPSYP-----TKSKPTTKPK 104
QY 247 ETTTINKOTSDGKEKETSAKETOSIEKTSADLAPTSIVLAKPPKAEYTTTGGPALTTP 306
DB 105 IT-----YPTTKAKPSYSPYKPKTKYPTTK-----PKLTY-----PPYKPK 143
QY 307 KKPPTTT--KPEASTPKPEPTTITKSAPTTP-----KEPATTTKSAPTTKKEPAP 357
DB 144 KPSYPTTKPKPSYSPSTTKTKTYSSTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 199
QY 358 TTTKEPA--PTTKPEAPPTTKPEAPTTKSAPTTKPEPA--PTTKPKPA--PTTKPEPA 411
DB 200 TYKAPTYKAKPTYSYTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSY 259
QY 412 -PTTKKEPT--PTTKPEAPPTTKPEAPTTKPEAPPAAPKPA--PTTKPEPA--PTTKKEP 465
DB 260 PPTTKAKPTTKAKPTTKAKPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 315
QY 466 A--PTTKEPS--PTTKPEPA--PTTKSAPTTKPEPA--PTTKSAPTTKPKPSPTTKKEP 519
DB 316 SYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 371
QY 520 APTTKKEAPPTTKPKPA--PTTKPEPA--PTTKPEAPTTTKKPAAPKPEAPPTTKKET 575
DB 372 SYPTTKAKPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 424
QY 576 APTTPKTLPTTPEKLAPTTEKPAPTTPELAPTTEBEPTTPEDEPAPPTTKAAAPNT 635
DB 425 YPST-YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 468
QY 636 KPEAPPTTKPEAPPTTKPE--PAPTTKETAPT--PKGTAPTTLKEPAPTTPKKAP 689
DB 469 PSYSAVYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 527
QY 690 KELAPTTTKETSTSDKPAPTTKGTAPTT--KPEAPPTTKPEAPPTTKGT--AP 742
DB 528 KTYPEPTTK-----PKISYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 573
QY 743 TTKLEPA--PTTKPKPA--KPELAPTTKG--PTSTSDKAPPTTKKATATTTKEPA--P 795
DB 574 TYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 629
QY 796 TTPKKA-PIT--PPTPPTSEVSTPTTKKE--PTTIKSPDESTPELSAPPTPK 846
DB 630 TYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 688
QY 847 ALENSPK-----EGGVPTT--KTPA-----ATKPEMTTAKOKTDEDLKTTPETTTAAPK 895
DB 689 AKPTNPSTYKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 747

```

```

QY 896 MKETATTEKTESKITATTTQVSTTTQDTPFKITLTKTTIAPKVTTKTITTE 955
DB 748 -TYKAKPTTPTPKAK-----PSTPPYKPKPKSPPTPKSKSPSSYKPKKKTTPPT- 798
QY 956 INKPEETAKPKDRATNSKATPKPKPTKAPKK--PTSTKPKTTPVRYKPKTTPTP- 1011
DB 799 --YKPLTTPPYK-----PKSPSPSKPKPKTTPSTYKLPSPPTPKSKTSYPT 848
QY 1012 -----RKMTSTMPPLNPTSR 1026
DB 849 YNKISYPSYSKAKTSYPPKPTNR 874

RESULT 6
ZAN_HUMAN STANDARD; PRT: 2700 AA.
ID ZAN_HUMAN 000218;
AC 09Y493; 000218;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, last sequence update)
DE 20-AUG-2001 (Rel. 40, last annotation update)
DE ZONADHESIN (FRAGMENT).
GN ZAN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE OF 1-2379 FROM N.A.
RX MEDLINE-99018118; PubMed-979793;
RA Glockner G., Scherer S., Schlettev R., Boright A., Weber J.,
RA Tsui L.C., Rosenthal A.;
RT "Large-scale sequencing of two regions in human chromosome 7q22:
RT analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci
RT reveals 17 genes.";
RL Genome Res. 8:1060-1073(1998).
RN [2]
RP SEQUENCE OF 2338-2700 FROM N.A.
RX TISSUE-Testis;
RX MEDLINE-97271566; PubMed-9126492;
RA Gao Z., Hartung T., Gathers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN).";
RL Genomics 41:119-122(1997).
CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATE SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -1- DOMAIN: THE VWFD DOMAINS 2 AND 3 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
CC -1- SIMILARITY: CONTAINS 3.5 MAM DOMAINS.
CC -1- SIMILARITY: CONTAINS 4.5 VWFD DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF053356; AAC78790.1; -
CC MIM: UB3191; AAC51208.1; -
CC DR EMBL: 602372; -
CC DR InterPro: IPR000561; EGF-like.
CC DR InterPro: IPR000998; MAM.

```

```

DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR002919; TIL.
DR InterPro: IPR003328; TILA.
DR InterPro: IPR001007; VWFC.
DR InterPro: IPR001846; Wvd.
DR Pfam: PF00629; MAM; 3.
DR Pfam: PF01825; TIL; 4.
DR Pfam: PF02345; TILA; 4.
DR Pfam: PF00094; wvd; 4.
DR PRINTS: PRO1217; PRICHEXTENS.
DR SMART: SM00137; MAM; 2.
DR SMART: SM00214; VMC; 1.
DR SMART: SM00011; VMC.def; 3.
DR SMART: SM00216; WVD; 3.
DR PROSITE: PS01186; EGF_2; 3.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS50060; MAM_2; 4.
KW Glycoprotein; Transmembrane; Cell adhesion; Repeat.
FT NON_TER 1 109 MAM 1.
FT DOMAIN 112 136 MAM 2 (PARTIAL).
FT DOMAIN 161 326 MAM 3.
FT DOMAIN 322 446 MAM 4.
FT DOMAIN 483 951 66 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
(MUCIN-LIKE DOMAIN).
FT DOMAIN 953 1065 VWFD 1 (PARTIAL).
FT DOMAIN 1066 1454 VWFD 2.
FT DOMAIN 1455 1861 VWFD 3.
FT DOMAIN 1862 2292 VWFD 4.
FT DOMAIN 2293 2684 VWFD 5.
FT DOMAIN ? ? EGF-LIKE.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1023 1023 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1618 1618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1737 1737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1832 1832 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1878 1878 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2136 2136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2505 2505 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 2374 2379 NNQKMA -> RAGPGP (IN REF. 1).
FT NON_TER 2700 2700
SQ SEQUENCE 2700 AA; 293013 MW; 80B60CC0B12277B1 CRC64;

Query Match 7.6%; Score 555.5; DB 1; Length 2700;
Best Local Similarity 32.6%; Pred. No. 9, 1e-15;
Matches 199; Conservative 64; Mismatches 240; Indels 107; Gaps 32;
QY 216 PINRPSLPNSDTSKETSITVKKETVETKETTNNKQTSIDGKKTSAKETOSIEKT 275
DB 445 PVKVLPELPVSPVSS---TGPESTTGLENTPTSTK-----KPTVSIKRP 487
QY 276 SAKDLAPTSVLAKPKPKATTKGVALTPPKPPTPKPEASTPKPKPT-----PTTI 330
DB 488 SVTTEKPT-----VPKKKPTIPEKPISTEK---PTIPEKNNMSEKPTIPEKPTIL 539
QY 331 KSAPTTKPEAPPTTKSAPTTPKEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKSAPT 389
DB 540 TEKTIPE-KPTIPEKPTISTEKPTVPEE--PTTPEETTYVEEPVPEKPSIPT 596
QY 390 TPKPAPPTTKKPAPTPKPEAPPTTKPEAPPTTKPEAPPTTKPEAPPTTKKPAAPKP 449
DB 597 --EKPSIPEK---PTISMETIISTEKPTICEKPIPIPEK---PTIPEKTIISPEK 647
QY 450 APTPKPEAPPTTKPEAPPTTKKPSPTTPKEAPPTTKSAPTTKPEAPPTTKSAPTTPK 509
DB 648 -PTTPTE-KPTIPEKPTISTEKPTIPEK-PTISPEKTLIPEKTLIPT---EKPTIPT 701
QY 510 EPSPTTKPEAPPTTKKPA-----PTTPKEAPPTTKPEAPPTTKKPAAPTAP 564
DB 702 EKPTISTEE--PTTPTEETIISTEKPSIPMEKPTIPEETTSVEETIISTEKTLITPM-- 757

```

QY 565 KEPAATPKETAPT-----TPKKLTPTTPEKLA-----PTTPEKAPATTPPEELATPTEE 614
 Db 758 EKPTISTEKPTIPKEPTISPEKLTIP-T-EKLTIPTEKPTIPLEETISTEKL--TIPTE 814
 QY 615 PTPTPPEAPATTTKAAAPNTPKAPATTTKEAPATTPPKAPATTPKEAPATTPPKGTAAP 674
 Db 815 KPTISPEKPTISTEK-----PTIPTE-KPTIPTE-----ETTISTEKLITPT 855
 QY 675 TLKEAPATTPKKAPKELATTTKEPTSTT-----SDKAPATTPGTAATTPKEAPATTT 728
 Db 856 --EKPTISPEKLTIPKEPTISTEKPTIPTEKLTIPTEKPTIPTEKPTIPTE--EKLALR 911
 QY 729 PKAPATTPGTAATTPKEAPATTPKKAPKELATTTKGTSTSDKAPATTPKET--A 786
 Db 912 PPHSPATA-TGLAALVMSPHAPSTPMTSV---ILGTTTSSSTGMSCP-PNARYESCAC 966
 QY 787 PTPPKPEAPT 796
 Db 967 PASCKSPRPS 976

RESULT 7
 EXTN_TOBAC
 ID EXTN_TOBAC STANDARD: PRT: 620 AA.
 AC P13983;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-MAR-1992 (Rel. 21, Last annotation update)
 DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
 GN HRPMT3.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; eusterids 1; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxId=4097;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. XANTHI; TISSUE=leaf;
 RX MEDLINE=90128263; PubMed=2612509;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich
 glycoprotein gene in lateral root initiation.";
 RL Genes Dev. 3:1639-1646(1989).
 CC -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
 THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
 MAIN ROOT.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
 CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 GLYCOSYLATED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X13885; CAA32090.1; -
 DR PIR: S06733; S06733.
 DR Repeat: Cell wall; Glycoprotein; Signal; Structural protein;
 KW Hydroxylation.
 FT SIGNAL 1 ?
 FT CHAIN ? 620
 FT REPEAT 70 73 EXTENSIN.
 FT REPEAT 148 151 H-A-P-P.
 FT DOMAIN 229 242 H-A-P-P.
 FT REPEAT 229 235 2 x 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 236 242 1.
 FT DOMAIN 205 620 2.
 FT DOMAIN 499 600 CONTAINS THE SER-PRO(4) REPEATS.
 3 x APPROXIMATE TANDEM REPEATS.

SO SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;
 Query Match 7.6%; Score 551; DB 1; Length 620;
 Best Local Similarity 27.2%; Pred. No. 4e-15;
 Matches 170; Conservative 64; Mismatches 304; Indels 88; Gaps 24;

QY 294 AETTK-GRALTP--KEPTPT-----PKEPASTTKE--PPTTIKSAPTTPKEAP 342
 Db 24 AEATLQYGVGLRPVTSQRPSSIGLSPSAPTTPPSRGHVSF--KAPAHNAVPP 80
 QY 343 TTTKAPPTTKEP-----APTTPKEAPT--TPKEAPATTTKEAPATTTKSAPTTPE 393
 Db 81 SHGHLRPSVGGPRPHRGHLRPSRNGFNPPSSVISPSSHPPSYGAPRPSHGHLRPSGGR 140
 QY 394 PAPTPPKKAPPT--TPKEAPATTPKEPTTPTPKEAPATTK--EPATTTKEAPATPK 448
 Db 141 P-----PSPSHGAPRSSGCHTTPKGNHPPSHRPSRPSRHGHPRPPTVAQRPPTIYS 193
 QY 449 PAPTPKEAPATTPKEAPATTTKEPSPTPK--EPAPTTKSAPTTTKEAPATTTKS-- 503
 Db 194 PDPV--QRPPT--YSPRPETHVQRPSPSRGHOQRPRTNRRAPRTNRRAPRTNOPSPL 249
 QY 504 ---APTTPKEPSPTTKEAPATTPKEAPATTPPKKAPATTPKEAPAT--TPKEAPATTK 557
 Db 250 RHLRPSRRQRPPTTSPRPAYASQSPPTYSPPRPYSPPRPYSPPRPYSPPRPAYSPSP 309
 QY 558 KPAPT--AKKEAPATTPKEAPATTPKKL-TPTT-----PELATTPTEKAPATTPREL 607
 Db 310 PPTPTPFSPPRPAYSPRPPTYLRLPSSPTYSPPRPYSPPRPYSPPRPYSPPRPPTYL 369
 QY 608 APTTPPEPTTTEBEAPATTPKAAAP--NTPEKAPATTPKEAPATTPKEAPATTPKETA 664
 Db 370 PPRPSSPPRPSPPRPYSPPRPYEOSPPRPAYSPRLAPRTYSPPPT--YSPRPYVQRP 427
 QY 665 PTPPKGAPATTTKEAPATTPKKAPKELATTTTKEPTSTSDKAPATTPKGAPATTPKEP 724
 Db 428 P-----LPRTYSPRPAYSPRPPTYSPPRPYSPPRPAYVQRPPTTYSPPRPAYSP 481
 QY 725 APTTPKEAPATTPKGAPATTTKEAPATTPKKAPKELATTTTGTSTSDKAPATTPKE 784
 Db 482 PPRSPYSPRPPOVORLPRTFSPPRRIRHLRPPHROG--RPTPTYGQRPSPPTSP 538
 QY 785 TAPTPKEAPAT--TPKKAPATTPETRPPTSEVSTPTTKEPTTIHKSPEDESTPELSAE 842
 Db 539 PPRQIHSPRRPHMQRTPTTYGQRPSPPTFSAPRPQIHSPRRPHQRPPTTYGQRP 598
 QY 843 PTPKALENSPKERGVPPTTKPATKP 868
 Db 599 PSP-----PTTYSPPSPPP 612

RESULT 8
 NFH_MOUSE
 ID NFH_MOUSE STANDARD: PRT: 1087 AA.
 AC P19246; Q61959;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
 GN NFH OR NFH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=89121513; PubMed=3220257;
 RA Julien J.-P., Cote F., Beaudet L., Sdky M., Flavell D., Grosveid F.,
 RT "Sequence and structure of the mouse gene coding for the largest
 neurofilament subunit.";

QY	310	TPPTPKPE-PASTPKEKPTPTTKSAPTTKEKPAAPTTTKSAPTTKEKPAAPTTTKKEAPATTKEKPAATTTP	368	
Db	599	TPSTPADSSAHSHTPSTPVDS---SAHSHTPSTPADSSAAGTSTPVDSASHT---PSTP	651	
QY	369	KE-PAATTTKEKAPTTTKSAPTTKE-PAATTPKKP-----APTTPKPEA---PTTP	415	
Db	652	ADSSAAGTSTPVDSASHTPSTPVDSASHSHTPSTPVDSASHGAPSTPADSSAAGTSTP	711	
QY	416	KEPTP-TPTEKAPTTKEKAPTTKEKPAAPPAKKPAATTPKEKPAAPTTKEPS	474	
Db	712	VDSASAGTSTPADSSASHTPSTPADSS-----AHSPTPADSSASHSPTPVDSAA	764	
QY	475	PTTPKEAPATTTKSAPTTKEKPAATTTKSAPTTKEKPSPTTKEKPAATTTKEKPA---	PT	530
Db	765	HGPTSTPADSSAHSHTPST---PADSSAAGTSTPVDSASHT---PSTPVDSASAGTSP	817	
QY	531	TP-KKPAATTPKEKAPTTTKEKPAATTTKKPAATPAKKEAPATTKEATPTPKL---	TP	585
Db	818	TPVDSASHTPSTPVDSASAGTSPSTPVDSASHSHTPSTPADSSASHT-PSTPADSSAAGTSP	876	
QY	586	TPP-ELKAPTTPEKAPTTPELAPTTPEE---PTTPTPEEPA---PTTP-KAAPNT	635	
Db	877	STPVDSASHSHTPSTPADSSAS---TPSPVDSASHSHTSTPADSSAAGTSTPVDSASHT	935	
QY	636	PKKPAATTPKEKAPTTTKEKPAATTPKATPTTPKGAATPTTKKPAATTTKKAAPKELAT	695	
Db	936	PSTPADSS---AHSPTPADSSASHT-PSTPADSS-----AHSPTPVDSASHT	983	
QY	656	TKKEPTSTTSKKAATTPKGAATPTTKE-PAATTPKEKAPTTTKGATTTKEKPAATTPK	754	
Db	984	-----PSTPADSSASHT-----PSTPADSSASHTSTPADSSASHSHTPSTPVDSASHSHTP	1033	
QY	755	KPAKELAPTTTKGPTSTSDKPAATTPK---ETAPTTKEKAPTTTKKAPTTPEPTP	810	
Db	1034	TP-----ADSSAAGTSTPADSSASHSHTPSTPVDSASHSHTPSTPADSS---AHSPTSP	1084	
QY	811	PTTSEVSTPTTKKPTTTHKSPDESTP-ELSAEPTPKALENSPKKPGVTTTPATPKPE	869	
Db	1085	DSSAH-STPSTPAD-SSAAGTTP-STPADSSASHTP-----STPAGSSAN	1125	
QY	870	MT	871	
Db	1126	GT	1127	

RESULT 10

CPN_DROME

ID	CPN_DROME	STANDARD:	PRT:	865 AA.
AC	002910;			
DT	01-OCM-1993 (Rel. 27	Created)		
DT	01-OCT-1993 (Rel. 27,	Last sequence update)		
DT	01-FEB-1994 (Rel. 28,	Last annotation update)		
DE	CALPHOTIN.			
GN	CPN OR CAP.			
OS	Drosophila melanogaster (Fruit fly).			
OS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
ON	NCBI_TaxID=7227;			
ON	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CANTON-S;			
RC	MEDLINE=93165729; PubMed=8094559;			
RA	Marlin J.H., Benzer S., Rudnicka M., Miller C.A.;			
RT	"Calphotin", a Drosophila photoreceptor cell calcium-binding protein.,"			
RL	Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CANTON-S;			
RC	MEDLINE=93165730; PubMed=8434015;			
RA	Ballinger D.G., Xue N., Harshman K.D.;			
RT	"A Drosophila photoreceptor cell-specific protein, calphotin, binds			
RT	calcium and contains a leucine zipper.,"			

```

RL  Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC  -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-STORAGE "SPONGE" TO
CC  REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
CC  OF CA+2 PER MOL OF PROTEIN.
CC  -1- SUBUNIT: HOMODIMER (PROBABLE).
CC  -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC  -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
CC  COMPOUND EYES AND OCELLI.
CC  -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
CC  DEVELOPMENT.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; L02111; AAA28405.1; -
DR  EMBL; L03080; AAA28420.1; -
DR  PIR; A47282; A47282;
DR  FlyBase; FBgn0010218; Cpn.
KW  Calcium-binding.
FT  CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
FT  CONFLICT 43 43 I -> T (IN REF. 2).
FT  CONFLICT 64 64 I -> V (IN REF. 2).
FT  CONFLICT 76 76 T -> A (IN REF. 2).
FT  CONFLICT 100 100 P -> PP (IN REF. 2).
FT  CONFLICT 126 127 VQ -> AP (IN REF. 2).
FT  CONFLICT 154 154 I -> V (IN REF. 2).
FT  CONFLICT 160 160 S -> T (IN REF. 2).
FT  CONFLICT 534 534 A -> E (IN REF. 2).
FT  CONFLICT 699 699 I -> T (IN REF. 2).
FT  CONFLICT 703 703 V -> L (IN REF. 2).
FT  CONFLICT 721 721 D -> E (IN REF. 2).
FT  CONFLICT 721 721 D -> E (IN REF. 2).
SQ  SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFC CMC64;

Query Match 7.3%; Score 530; DB 1; Length 865;
Best Local Similarity 26.2%; Pred. No. 3,4e-14;
Matches 233; Conservative 73; Mismatches 433; Indels 152; Gaps 41;

```

```

DB  414 V-----LPPVAAEPVPAVVAEETPETPAPASAVTIALDIPEVAPVIAAPSDAPAPASPA 469
OY  692 LAPTTKEPTSTSDKPAPTPKGTAPTP-----KEPAPTPKPE--APTTPKGTAPT 743
DB  470 AAPVSTPPT--TTSVDETTAPPAVAVPEPDIIVSVSEALIEFPVAPVAVTTEVAVADV 527
OY  744 TLKEPA-----PTTPKKPAPELAPTTT-----KGPSTTSDDKPAPTT--PKETAPPT 789
DB  528 APPENAADLIEVEPPAPPIDLEQTTSVAVAASTSS--PIPETSLPPNEAVASP 585
OY  790 KEPAPTTPKRPAPTPPETPTTSEVSTPTTKEPTTIKSPD-----ESTPEL 839
DB  586 EVAVAPITAPPEIP-----EPEPSLATPEPIPEVAPVIOEADAVAEVETETSTIPET 641
OY  840 SAEPPPKALENSPKREGVPTTKTPAATKPEMTTAKOTTERDRTTP--ETTTAPKMT 897
DB  642 TVE-PPEVAAEKVIDPAL--TEAPVTTIODEPVANINOCAPATEI--TTPAVEIVTAAAEVS 697
OY  898 KETATTEKTESKITAVTTTOST-----TQDTTPFKITTLKTTTLA--PKYTTTKT 950
DB  698 DIAIPVIDPPVQELIYAVAEIPETDKPAVEIVEOSTTIPTEAPVPEVSKYAPVISEAPAA 757
OY  951 ITTTEIMNKEETAKPKDRATNSKATTPKPKPKAPKPKSTKPKPKTPMPVRKPKTTP 1010
DB  758 EAPITAGDMDPNTSVGISEVPTIAEKPEVEVPIS--EIPQSSSPS--DSVPAKITPL 813
OY  1011 PRKMTSTPELNPSTRIAEAM--LQTTTPNQTNSKLVEVN-----PKS 1053
DB  814 LRDLQTTDVSILAIATIDAIIGELKDKQARNQOVMDRICEIKILGPKS 864

RESULT 11
PPL MYTCO STANDARD; PRT; 872 AA.
AC  Q25434;
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  20-AUG-2001 (Rel. 40, Last annotation update)
DE  ADHESIVE PLAQUE MATRIX PROTEIN PRECURSOR (FOOT PROTEIN 1) (MCP1).
GN  PPL.
OS  Mytilus coruscus (Sea mussel).
OC  Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;
OC  Mytiloidae; Mytilidae; Mytilus.
OX  NCBI_TaxID=42192;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  MEDLINE=96394686; PubMed=8798340;
RX  Inoue K., Takeuchi Y., Takeyama S., Yamahara E., Yamazaki F., Odo S.,
RA  Harayama S.;
RA  "Adhesive protein cDNA sequence of the mussel Mytilus coruscus and
RT  its evolutionary implications.";
RL  J. Mol. Evol. 43:348-356(1996).
CC  -1- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS
CC  PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GUE. THE MUSSEL'S
CC  ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A
CC  FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.
CC  -1- SUBCELLULAR LOCATION: SECRETED.
CC  -1- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.
CC  -1- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.
CC  -1- PTM: TYROSINE (THUS PRODUCING DOPA - 3,4-DIHYDROXYHENTYLAAMINE).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D63777; BAA09850.1; -
DR  InterPro; IPR002964; Adhesive_plaq.

```

DR InterPro: IPR002965; P_fich_extensn.
 DR PRINTS: PRO1216; ADHESIVEI.
 DR PRINTS: PRO1217; PRICHEXTENSN.
 KW SIGNAL; Repeat; Hydroxylation.
 FT CHAIN 1 20 POTENTIAL.
 FT CHAIN 21 872 ADHESIVE PLAQUE MATRIX PROTEIN.
 FT DOMAIN 1 41 NONREPEAT LINKER.
 FT DOMAIN 124 872 TANDDEM REPEATS OF Y-K-[PS]-K-[IP]-[ST]-Y-
 FT DOMAIN 184 192 P-[PST]-[ST].
 FT DOMAIN 213 NONAPEPTIDE 1.
 FT SEQUENCE 872 AA; 101677 MW; 98CC70D7C75F3C4 CRC64;

Query Match 7.1%; Score 518.5; DB 1; Length 872;
 Best Local Similarity 29.2%; Pred. No. 9.3e-14;
 Matches 276; Conservative 80; Mismatches 407; Indels 181; Gaps 54;

QY 177 VVDEAGSGLDNDFVYTPDST--TOH-----NKSTSPKITTAKPIPRESLPNS 227
 DB 29 YGSAVSGSAGAYK-TLPGSHPGSKHVPYKPMNK1PT-PYI--SKKSPAPYKPKGY 84
 QY 228 DTSKETSILVNKEETVETKETTNNKQSTDGKEKTSKETSOSIEKTSADKLAPTSKVL 287
 DB 85 YPTKRYOPTYGSKTYNPIYKPIAKKLSSYKAITTYPAVAKTSYPSYK-----HKIT 139
 QY 288 AKPTPKAETTKGPAITTPKEPT--PTTPKEPASTTPKEPPTTKSAPTTKEPAPTTP 346
 DB 140 YPTPK-----PKITTP--PYTKOKSPSYKPKITTYPTPK-----PKITTPPTPK 185
 QY 347 SAFT--TPKEPAPT--TPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPT--PK 401
 DB 186 RKPSTYTPKPKATYPTPKITP---PYTKKPSYK-----PYKPKITTYPTPKPKI 236
 QY 402 PAPTPKEPAPT--PKETPTT--PKPAPTTPKEPAPTTPKEPAPT--APKPAPTTP 454
 DB 237 SYPSYTPKASYSSYSKSKITTYPTPKKISYPTPKPSYP-----PYTKPVTPPT 292
 QY 455 KEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPT--TPKEPAPTTPSAPTTPK 509
 DB 293 K-PKPSYPTPKKITPTPT--YKPKPSYPTPKOKPSYPTPKSSYSKSKITTP 350
 QY 510 EPSPTTPKEPAPTTPKEPAPTTPKAP--APTPK-----EPAPTTPKEPAPTTPK 559
 DB 351 TYKPKITTYPTPKKPSYPSYKPKKTSYPTPKKITTYPTPKPSYPSYKPKITTP 410
 QY 560 APAPKEPAPTTPKEPAPT--TPPKKLPTTPKEKIAPTTPKPAPTTPPEELAPTPPEE 614
 DB 411 T-YKPKISYPTPKPKASYSYSKSKITTYPTPKKISYPTPKPKPSYPTPKKITTP 469
 QY 615 PT-PTTPKEPAPTTPKAAAPNTPK--EPAPTTPKEPAPTTP--KEPAPT--PKETAP 665
 DB 470 PTYKPKPSYPTPKKITTYPTPKKPSYPTPKOKPSYPTPKSSYSKSKITTP 529
 QY 666 TT--PKGTAFTTPKEPAPTTPKPAKELAPTTKEPTSTDSKAPAT--TPKGAFTTPK 722
 DB 530 PTYKPKITTYPTPK-----PKPSYPSYKPKITTYPTPKKIRYPTPKPKASYPPTPK 583
 QY 723 EPAPTTPKEPAPTTPKGAFTTPKEPAPTTP--KKPAKELAPTTTGPSTISDKPA 778
 DB 584 -----PKITTPPTPK--PKPSYPTPKOKPSYPTPKSSYSKSKITTP 628
 QY 779 PTT--PKETAPTPKEPAPTTPK--KPAPTPEP-----PPPTSEVSTTPTKPEPT 827
 DB 629 PPTPKKITTYPTPK--PKPSYPSYKPKITTYPTPKPKKSPQAVKSGSPSYQPKKT 687
 QY 828 IHKSPDESTPELSAETTPALENSPEKPEVPTTKPAATKEMTTAADKT-----TEDEL 883
 DB 688 Y--PPSYKPKKITTYPTPK-----PKISYPTPK--TKPSYPSYKSKITTPPTPK 735
 QY 884 RTTPEPTTAAPKMKETATTEKTESKITATTTQVSTTTODTTPFKITTLAKTTTAPK 943
 DB 736 ISYPTPKKAP-----SYPTPKPKPSYASSYKPKIRPPTPKPKPSYASSYK-----PK 785

QY 944 VTTTKKITTITTEIMNKPEETAKPKDRATNSKATTPKPKP---KPTKAPKRP-TSTKPK-T 998
 DB 786 I-----RPPPTPKPKP-----SYASSYKPKIRPPTPKPSYASSYKPKIT 827
 QY 999 MPVKKPKPT--PTPKMTSTMEELNPTSRIDAMQOTTTRPQ 1040
 DB 828 YPTPKKISYPTPKKITTYPTPKKISYPSYKPKISYPSQ 871

RESULT 12
 ID NFN_RAT STANDARD: PRT: 831 AA.
 AC P16884; 063368;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).
 GN NFN OR NFN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=101116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain:
 RX MEDLINE=89065087; PubMed=3143606;
 RA Breen K.C., Robinson P.A., Wion D., Anderson B.H.;
 RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
 RT Identification of putative phosphorylation sites.";
 RL FEBS Lett. 241:213-218(1988).
 RN [2]
 RP SEQUENCE OF 37-831 FROM N.A.
 RX MEDLINE=88309090; PubMed=2457365;
 RA Dautigny A., Phan-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,
 RA Jolles P.;
 RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
 RT in situ detection.";
 RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
 RN [3]
 RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
 RX MEDLINE=87080760; PubMed=2878828;
 RA Robinson P.A., Wion D., Anderson B.H.;
 RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
 RT (NF-H).";
 RL FEBS Lett. 209:203-205(1986).
 RN [4]
 RP SEQUENCE OF 318-831 FROM N.A.
 RX MEDLINE=89184647; PubMed=2928342;
 RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
 RA Smolowitz M., Carroll Z., Emanuel B.S., Breilner J., Rubin L.;
 RT "Cloning of a cDNA encoding the rat high molecular weight
 RT neurofilament peptide (NF-H): developmental and tissue expression in
 RT the rat, and mapping of its human homologue to chromosomes 1 and
 RT 22.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFN IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFN RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
 CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.

DR	PROSITE: PS00025; P_TREPOIL; 6.
KW	Repeat: Amphibian skin; Glycoprotein; Alternative splicing.
FT	NON_TER 1 1
FT	DOMAIN 81 144
FT	REPEAT 81 88
FT	REPEAT 89 96
FT	REPEAT 97 104
FT	REPEAT 105 112
FT	REPEAT 113 120
FT	REPEAT 121 128
FT	REPEAT 129 136
FT	REPEAT 137 144
FT	DOMAIN 161 202
FT	DOMAIN 218 301
FT	REPEAT 218 224
FT	REPEAT 225 239
FT	REPEAT 240 249
FT	REPEAT 250 259
FT	REPEAT 260 275
FT	REPEAT 276 287
FT	REPEAT 288 294
FT	REPEAT 295 301
FT	DOMAIN 306 347
FT	DOMAIN 353 394
FT	DOMAIN 402 522
FT	REPEAT 402 411
FT	REPEAT 412 419
FT	REPEAT 420 431
FT	REPEAT 432 443
FT	REPEAT 444 453
FT	REPEAT 454 460
FT	REPEAT 461 472
FT	REPEAT 473 479
FT	REPEAT 480 491
FT	REPEAT 492 498
FT	REPEAT 499 515
FT	REPEAT 516 522
FT	DOMAIN 525 566
FT	DOMAIN 572 613
FT	DOMAIN 620 661
FT	DISULFID 162 186
FT	DISULFID 172 187
FT	DISULFID 182 199
FT	DISULFID 307 333
FT	DISULFID 317 332
FT	DISULFID 327 344
FT	DISULFID 354 380
FT	DISULFID 364 379
FT	DISULFID 374 391
FT	DISULFID 526 552
FT	DISULFID 536 551
FT	DISULFID 546 563
FT	DISULFID 573 599
FT	DISULFID 583 598
FT	DISULFID 593 610
FT	DISULFID 621 647
FT	DISULFID 631 646
FT	VARIANT 641 658
FT	VARIANT 276 276
FT	VARIANT 354 354
FT	VARIANT 415 415
SO	SEQUENCE 662 AA; 67774 MW; F0B5277E1ED2FD40 CRC64;

```

Query Match      7.0%   Score 509.5; DB: 1; Length 662;
Best Local Similarity 28.1%   Pred. No. 116e-13;
Matches 233; Conservative 48; Mismatches 276; Indels 247; Caps 30.

OY 250 TTNRKQSTQCKETKSTAFQETQSIETKTSKDLAPTSKVLAKTTPRAETTTGCPALTPKPK 309
      ||      ||      ||      ||      ||      ||      ||      ||      ||
      3  TTAAVAAGAAGKDTTAAAEQSAALAEKTA-----AGEVAPPT--AAVAANGDEAIT----- 51

```

QY	310	TTTTPKKEASTPKKEPPTTTT	KSAPTTKKEAPATTTK	KSAPTTKKEAPATTTKKEAPATTPK	36
Db	52	-----	AAATAAETTAAGAEPTTTT	ATTAAGAKPTTAATAAPTAAAGATTAT	10
QY	370	EPAPTTTKEAPATTTKS	-APTTPKEAPATTPKKPAPTTKEPATTKEP	-APTTPKEAPATTTKS	42
Db	104	GKAPATAAAPTAAASKATTT	AAATHTSIAAAAAPTAAASAKSERSTSSSEEHCH		16
QY	428	--PTTKPAPT---	TPKEAPATAKKKPAPTTPK---	EPAPTTKKEAPATTKE	47
Db	164	VKPSKRMCGSKGTTKKO---	CKKKNCCEDPKGHGIGHCFHKKRPGHSHHEHTTTTK-		21
QY	473	PSPTPKPEAPATTKSAPTTPKEAPATTKSAPTTPKE	SPPTTKESPTTPKEAPATTKEAPATTPK		53
Db	219	-----	APTITQATTTTT	-----PTTKKAPT---	24
QY	533	KKAPATTKEAPATTKEAPATTTKKPAPTAKEAPATT	PKETAPTTPKLTPTPEKLA		59
Db	246	-----	TTTTTKATPTTT	-----TTTKATTTTT	27
QY	593	PTTPKEAPATTPELAPTTPEEPPTTPKEAPATTPK	KAAPTTPKAAPTKEAPATTKEAPATTPK		65
Db	271	-----	TTTTTKATTTPTTTTTTT	-----	28
QY	653	KEAPATTKEAPATTGAPTTLKEAPATTKKPAPKELA	PTTTKEPTSTTSOKPAPT		71
Db	290	-----	TTTTTKATTTTTTSGECKMEPSK	-----REDCGSGITESOCR	32
QY	713	PKG----	TAPTT-----PKAPATTKEAPATTGKAPTTLKEAPATTKKPAPKELA		76
Db	329	TGCGCDDSIPTKNCFTLSO	ADCKVPESSORDCGRGIT----	ADOCROKMCCEPSS	38
QY	763	PTTTGCPSTSDKAPATTPKETAPTTKEAPATTPKKP	APATTPEPTPEPTSEVSTPTT		82
Db	385	ISGTTWCFSYSQVAA--	TKTTTTPTTTPTTTTTKATTTTTT	-----TTTTTPTTT	43
QY	823	KEPTTHKSPEDEPSELAE	PKALENSREPGVPTTKPAATKPEMTTANKDTERD		88
Db	437	TTTTT	-----TTKA-----	TTTTPTTTPTTTTTT	46
QY	883	LPTPEPTTAAPKMTKETATTTTEKTIATTTT	QVSTTODTPEFKITLTKTLAP		94
Db	464	TTTTPTTTTTTT	-----TTTTTKAT-----	TTTTTTTTTTTTTKATTTTTTTTTTT	51
QY	943	KVTTTKKTTTTTELMNKPPEELAKRDRATNSKATTP	PKPTAKPKKPTSTKRPK----	T	98
Db	511	TTTTTKATTTTTSECKME--	PSKRADCGPGCITESOCRSGCCPDSIPTKMCFTS		56
QY	999	MPRVAKPKPTPTTPR	1012		
Db	567	LPOVADCKVAPSSR	580		
RESULT 14					
RBP1_HUMAN					
AC	P24928	STANDARD	PRT	1970	AA.
DT	01-MAR-1992	(Rel. 21, Created)			
DT	01-MAR-1992	(Rel. 21, last sequence update)			
DT	20-AUG-2001	(Rel. 40, last annotation update)			
DE	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6) (RBP1).				
GN	POLR2A.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_Taxid=9606;				
OX	NCBI				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92178992; PubMed=1542581;				
RT	Winterlith M., Acker J., Vitale S., Vigneron M., Keding C.;				
RT	Complete sequence of the human RNA polymerase II largest subunit.				
TL	Nucleic Acids Res. 20:910-910(1992).				

DR PIR: S00979; QFHUH.
 DR MIM: 162230; -
 DR InterPro: IPR001664; IF.
 DR Pfam: PF00038; filament; 1.
 DR PROSITE: PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation.
 FT DOMAIN 1 100 HEAD.
 FT DOMAIN 101 413 ROD.
 FT DOMAIN 414 1020 TAIL.
 FT DOMAIN 101 132 COIL 1A.
 FT DOMAIN 133 145 LINKER 1.
 FT DOMAIN 146 244 COIL 1B.
 FT DOMAIN 245 266 LINKER 12.
 FT DOMAIN 267 288 COIL 2A.
 FT DOMAIN 289 292 LINKER 2.
 FT DOMAIN 293 413 COIL 2B.
 SQ SEQUENCE 1020 AA; 111780 MW; 1177C9DCB3DCFD1D4 CRC64;

Query Match Best Local Similarity 26.7%; Score 502; DB 1; Length 1020;
 Matches 209; Conservative 81; Mismatches 356; Indels 136; Gaps 35;

OY 102 KTKRYESE--EITEHSVS--ENQSSSSSSSSSTIWKIKSKNSANRELQKRLK 158
 DB 336 KSTKDSLEGRSELEDRHQADIASYOEAIQDLAEIRNTKMEM-----AAQLREYQDLIN 390
 OY 159 VR-----DNKNRTRKKRPKPPVVDAGSGLDNGDFKVTTPDTSTQHNK 204
 DB 391 VKMALDIEIAAYRKLEGECHRGFPD-----FSLPEGLPKI--PSVST--HIK 437
 OY 205 VSTSPKITAAPINPRPSLPPNSDTSKETSIVNKETIVETKETTNNKQSTDGKE-KT 263
 DB 438 VKSEKIKVVE-----KSEKTVIVER-----QTEIQVTEVEVEEKEKEAKE 480
 OY 264 TSAKETQSIKTSANDLAPTSKVLAKPTPKAETTKGPAITTPKEPTTPKKEPASTTPK 323
 DB 481 EEGKEEGEGEEEAEG-----GEEETKSP-----PAEEAASPEKEAKSPVKE 522
 OY 324 EPTPTTISAPPTTKEPAPTTPKTSAPPTTPKEPAPTTPKKEPAPTTPKKEPAPTTP 383
 DB 523 EAKSPAEEAKSPKEAKSPAEEVKS--PEKAKSPA-----KEEAKSPPEAKSPKEEAKSPA 577
 OY 384 TKSAPPTTKEPAPTTPKPPAPTTPKEPAPTTPKKEPTTPTPKEPAPTTPKKEPAPTTP 443
 DB 578 VKS--PEKAKSPAEEAKSPA-----EAKSPAEEAKSPA--EAKSPAEEAKSPA 631
 OY 444 TAPKKPAPTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTP 503
 DB 632 VKSPAEEAKSPKEAKSPAEEAKSPAEEAKSPAEEAKSPA--PEKAKSPAEEAKSPA 690
 OY 504 APTTPKEPSPPTTTPKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTPKKEPAPTTP 563
 DB 691 -PEKAKSP-----VKEEAKSPAEEAKSPAEEAKSPAEEAKSPA--KEEAKTPPEKAKSPV- 740
 OY 564 PKPAPTTPKETAPTTPKKTLPPTPEKLAPTPE--KPAVTPPEEL--APTTPPEEPTTP 620
 DB 741 -KEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPAEEAKSPA 799
 OY 621 EEPAPTTPKAAAPNTPEPAPTTP-----KEPAPTTPKEPAPTTPKETAAPTTPKGTAPPTL 676
 DB 800 KSPLEKADAKAPKEIKPEKEEVKSPVKEEKEPOEVKVEPPKRAEEKAPATPK----- 852
 OY 677 KEPAPTTPKPAKELAPTTTKEPTSTSDKPAPTTPKGTAPPTTPKEPAPTTPKKEPAPTTP 736
 DB 853 TEEKKDSKKEEAPKEAP-----KPVEEKKPEPAVEKPEKSKVEAKKEAEKKKVP---T 905
 OY 737 PKGTAP--TTLKEPAPTTPKPAKELAPTTTKEPTSTSDKPA--PTTPKETAPTTPKEP 793
 DB 906 PEKEAAKVEVKEA-----KPKETEVAKKEPDADAKAPEKSKPAEKEEAAP--EKKOT 957
 OY 794 APTTPKPAAPTTPPEPTTPTEV--STPTTKEPTTIHKSPEDE--STPELSAEPTPKAL 848

DB 958 KEKAKK-----PEEKPKTEAKAKEDDKTLSKEPSKPKAEKAEKSSSTDOKDSKPEKAT 1012
 OY 849 EN 850
 DB 1013 ED 1014

Search completed: April 26, 2002, 16:32:21
 Job time: 571 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:25:14 ; Search time 61.21 Seconds
(without alignments)
501.094 Million cell updates/sec

Title: AA2
Sequence: 7276
1 MAWKLPILYLLLLSVEFIQ.....ARATRRSGQLSKVYNCP 1363

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	522	7.2	805	Sequence 4, Appl
2	513	7.1	1837	Sequence 5, Appl
3	506.5	7.0	5202236-25	Patent No. 5202236
4	499.5	6.9	1721	Sequence 5, Appl
5	498.5	6.9	786	Sequence 3, Appl
6	498.5	6.9	1721	Sequence 6, Appl
7	488.5	6.7	826	Sequence 2, Appl
8	488.5	6.7	826	Sequence 2, Appl
9	476.5	6.5	5202236-13	Patent No. 5202236
10	452	6.2	960	Sequence 5, Appl
11	425.5	5.8	1185	Sequence 23, Appl
12	424.5	5.8	1867	Sequence 5, Appl
13	424.5	5.8	1867	Sequence 5, Appl
14	424.5	5.8	2035	Sequence 2, Appl
15	424.5	5.8	2035	Sequence 2, Appl
16	419.5	5.8	2476	Sequence 2, Appl
17	417	5.7	829	Sequence 137, Appl
18	417	5.7	829	Sequence 53, Appl
19	417	5.7	837	Sequence 68, Appl
20	417	5.7	837	Sequence 103, Appl
21	417	5.7	837	Sequence 103, Appl
22	417	5.7	837	Sequence 103, Appl
23	417	5.7	837	Sequence 103, Appl
24	417	5.7	837	Sequence 103, Appl
25	413.5	5.7	907	Sequence 50, Appl
26	413.5	5.7	907	Sequence 19, Appl
27	404.5	5.6	408	Sequence 65, Appl

28	404.5	5.6	408	US-08-475-411A-65	Sequence 65, Appl
29	404.5	5.6	408	US-08-478-029A-65	Sequence 65, Appl
30	398.5	5.5	682	US-08-642-255-126	Sequence 126, Appl
31	398.5	5.5	682	US-08-397-633A-36	Sequence 36, Appl
32	387.5	5.3	1537	US-08-325-263A-2	Sequence 2, Appl
33	383.5	5.3	1231	US-08-904-263A-4	Sequence 4, Appl
34	381	5.2	1848	US-08-296-791-6	Sequence 6, Appl
35	381	5.2	1848	PCT-US85-10661A-6	Sequence 6, Appl
36	376	5.2	960	US-09-219-849-6	Sequence 8, Appl
37	373	5.1	1786	US-08-973-462-8	Sequence 8, Appl
38	371	5.1	761	US-08-707-237A-84	Sequence 114, Appl
39	371	5.1	762	US-08-642-255-114	Sequence 26, Appl
40	370.5	5.1	1064	US-08-642-255-62	Sequence 62, Appl
41	370.5	5.1	1187	US-08-320-558-28	Sequence 28, Appl
42	370.5	5.1	1187	US-08-545-860D-28	Sequence 28, Appl
43	370.5	5.1	1187	PCT-US94-04456-28	Sequence 28, Appl
44	370.5	5.1	1210	US-08-320-559-26	Sequence 26, Appl
45	370.5	5.1	1210	US-08-320-559-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-09-103-429A-4
Sequence 4, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
APPLICANT: Wang, Ping
TITLE OF INVENTION: A NO. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: CDNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Brown, Plinnel & Michaels, P.C.
STREET: 118 No. 6187558th Tloga
CITY: Ithaca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-3628
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 805 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
US-09-103-429A-4

Query Match 7.2% Score 522; DB 4; Length 805;
Best Local Similarity 27.8%; Pred. No. 9.2e-26;


```

Qy 535 ----- 534
Db 570 MIPNDTHVRFRKVKOVGNATISVRCGKAGKLEFPDRSLDTIPPVAGHNSCIIIVGS 629
Qy 535 -----PAPTPKE-----PAPTPKEPAPTPTK 558
Db 630 GGGGHHVSPGSKDVSLISAPICELFENVCYDCTAKGAIHSGYQISADVTTTAK 689
Qy 559 PAPAPKEPAPTPKEPAPTPKLPPTPEKLAFTPEPAPTPKEPAPTPPEELAPTPPEEPT 618
Db 690 PTTTT--TGAPGQPTTTTSSPSKPTTTTAKTTTT--TINPTTTTQKPTT 741
Qy 619 TPEPAPTPKEAARNTPKEPAPTPKEPAPTPKEPAPTPKEPAPTPKSTAPPTLKE 678
Db 742 T-----T--KVGPSPRIATTTTILKPIVTTTTKATTTTTPPT--TTTKR 787
Qy 679 PAPTPKPAKP--ELAPTTKEPTTS-----TSDKAPTPPG--TAP 718
Db 788 DEMTTTTPLPDIGDIEIRPIEKMLDKYTMIIYDNGSLLDSDDEIPGSOAQIAD 847
Qy 719 TTPKEPA-----PTTPKEPAPTPK----- 738
Db 848 TSLMFPVQTHKSTGLPDIIDVGLPDPKSGNLVHPYTNQMSGLSVSLAKNLTVDTE 907
Qy 739 --GTAPPTLK--BPAPTPKPAKELAPTT-----KAPTSTSDK 777
Db 908 TGLPDLDTLGLPDLVSLIPNPETGELFDISDEINNGTAGIVSGISASESILSQS 967
Qy 778 APPTPE-----TAPTPKEPAPTPKPAKPTTPPT--PPTTSVSTPT 821
Db 968 AFDIDATNMVGEFGLNIPATGVMIPOSLSGEQTPSPISDIOGIIPT--EVALANA 1024
Qy 822 TREPTTIHKSPESTPE-----LSAPTPKALENSPKEBGP 858
Db 1025 DKFKLSIPSPVESPEKDQKIDISIELMYDIESGRLIGOVSKRPIGSIAGDLN---P 1080
Qy 859 TTKTPAAKPEKTTAKDKTERDLRTPETTAAPKMTKETA--TTEKTESKITA 914
Db 1081 IMKPTQY-----DSVTGKPI--DPTTGLP--FNPTGHLIMPNTNNNTDSSFAG 1126
Qy 915 TTTQVSTTTODTPEKTTITLKTTLAPK--VTTTKTTTITTEIMNKPETAPKPKRAT- 971
Db 1127 AVKAVSGKITDVAVYGLPDELTGL--PKDFVSDIPNPSSTTGLVD--PSTGKPINNYTA 1183
Qy 972 -----NSKATTPPKQKPTKAPKPKSTKPKTKPMRPAKPKTTTPPKKMS 1016
Db 1184 GIVSGKRGLPRIEDENGNLFPSTKLPIDGNQLVNENSTIVSGSTGSKPKPIVNA 1243
Qy 1017 ---TMPE-----LNPTSRIAEAMLTTRPNQTPNSKLVEVNEKSEADAG 1058
Db 1244 GGGVVPDEAKDQADKDKGLIVPPTNSINKDVTNTQYSNTTGNL---INP--ETGKV 1297
Qy 1059 AEGETPHMLRPHVEMPEVTPD-----MDYLRVNPNGIILN----- 1096
Db 1298 IPGSLPGLSVNPSFNTPQOIDEITGKRVDTVGLPYDPSTGELIIDPATRLPIGSAVAGDE 1357
Qy 1097 MLSETETICKG-----PVDGLTTLNGLTV 1122
Db 1358 ILTEVLNITTTDEVYGLDLELTGLPRDPVPSGLPQLPNTGLV 1398

RESULT 3
520236-25
; Patent No. 5202236
; APLICANT: MAUGH, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
; SUSAN L.; MCCANDLIS, RUSS; WEI, TENA; FILPULA, DAVID
; TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
; PROTEIN
; NUMBER OF SEQUENCES: 39
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/528,762
; FILING DATE: 25-MAY-1990
; APPLICATION NUMBER: 82,456

```

```

; FILING DATE: 07-AUG-1987
; APPLICATION NUMBER: 933,945
; FILING DATE: 24-NOV-1986
; APPLICATION NUMBER: 650,128
; FILING DATE: 13-SEP-1984
; SEQ ID NO: 25
; LENGTH: 744
520236-25

Query Match
Best Local Similarity 29.9%; Score 506.5; DB 6; Length 744;
Matches 239; Conservative 78; Mismatches 308; Indels 175; Gaps 50.

```

```

Qy 209 KPTTAKPIMPRLPSNDSTKETSGLYVAKETVTEKTTTNNQTSYDGEKETSKE 268
Db 24 PKMTYPTTKPKPSYP-----YKSKPTV-----KPKIT----- 53
Qy 269 TQSLKTSANDLAPTSKVLAKPTPKAETTKGALTTPEPTTPKE-----PASTPKE 324
Db 54 -----YPTPK--AKPS-----YPTPKPKKTPPTPKKLTYPPTPKKP 92
Qy 325 PPTTIKSAPT--TPKEPAPTTKSAPTTPKEPAPTTKEPAPTTKEPAPTTKEPA--P 381
Db 93 SYPTTKSKPTTKPKLTYPPTKAKPSYPTPKPKKTYPPPT--YKPKLTYPPTKPKASYP 151
Qy 382 TTTKSAPTTPKEPAPTTTPKPAPT--TPKEPAPTT--PKBPTTPTPKEPAPTTKEP-- 433
Db 152 PTKPKPSYP--PSYTKTKTYPPTPKKLTYPPTKPKPSYPTPKPKKTYPPPKLT 209
Qy 434 -APTPKEPA--PTAKKPA--PTTPKEPA--PTTPKEPS--PTTPKEPA 482
Db 210 YPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 269
Qy 483 -PTTKSAPTTPKEPAPTTTKSAPTTPKEPSPTTKKEPA--PTTPKEPA--PTTPK 535
Db 270 YPTTKAKPSY--YPSYTKAKPTYP--PTTKAKPSYPTTKAKPSYPTTKAKPSY 322
Qy 536 A--PTTPKEPA--PTTPKEPA--PTTKKPA--PTAKKEPAPTTPEAPTTPEKLTPT 587
Db 323 SYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPS 382
Qy 588 PEKLAFTPE--KPAPTPPEELAPTTPEEPT--PTTBEPA--PTTPKA--AAPNT--P 636
Db 383 TYKAKPSYPTTKAKPSY-----PTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 438
Qy 637 KEPAFTTPKEPAPTT--KEPAFTTPKEPAPTT--PKGTAPPTLKEPAPTTPKPAKP 690
Db 439 SYPTTKAKISYPTTKAKPSYPTTKAKSSYPTTKAKSSYPTTKAKSSYPTTKAKP 498
Qy 691 ELAPTTKEPTSTSDKAPPTTPGTAFTTPKEPAPTTPEPAPTTPEKSTAPPTLKEPA- 749
Db 499 SYPTTKAKPSYPTTKAKPSY--ARPSY--PTTKAKPS-----YPTTKAKPSY 546
Qy 750 -PTTPKPAPELAPTTTKGPT--STSDKAPPTPEKETAFTTPKEPAPTTPKPAKP--PT 804
Db 547 PPTTKAKPTPK--AKPTKAKPTPKSTYTKAKPSYPTTKAKPSY--PTTKAKPSYPT 601
Qy 805 ---TPETPPTTSEVSTPTTKE-----PTTIKSPD-----ESTPELSAAPT--PVALENS 851
Db 602 YKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 661
Qy 852 PKBGPVPTTKPA-----TKPEMTTIAKDKTTEBDLTTPETTTAP-----KMT 897
Db 662 PSYP--PTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKPSYPTTKAKP 719
Qy 898 KETATTEKTTESKINATTT 917
Db 720 YKAKPTYPSTYKAKPTYPST 739

RESULT 4
US-08-700-651-5

```

```

: Sequence 5, Application US/08700651B
: Patent No. 6015882
: GENERAL INFORMATION:
: APPLICANT: PETERSEN, CAROLYN
: APPLICANT: LEECH, JAMES
: APPLICANT: NELSON, RICHARD, C.
: APPLICANT: GUY, JIRI
: TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
: TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF CRYPTOSPORIDIUM PARVUM
: TITLE OF INVENTION: INFECTIONS
: TITLE OF INVENTION: INFECTIONS
: FILE REFERENCE: 480.19-4(HV)
: CURRENT APPLICATION NUMBER: US/08/700,651B
: CURRENT FILING DATE: 1997-08-14
: EARLIER APPLICATION NUMBER: 08/415,751
: EARLIER FILING DATE: 1995-04-03
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: Patent Ver. 2.0
: SEQ ID NO 5
: LENGTH: 1721
: TYPE: PRT
: ORGANISM: Cryptosporidium parvum
: US-08-700-651-5

Query Match      6.9%; Score 499.5; DB 3; Length 1721;
Best Local Similarity 21.8%; Pred. No. 6.2e-24;
Matches 300; Conservative 94; Mismatches 510; Indels 469; Gaps 44;

OY 90 STTKRSPKPNKKTKTKVIESEITEHVSNSQSSSSSSSSSTTWIKS----- 143
DB 38 STSGCRKQBNKQAT-----ISSRSCGMKQGYSDSSGFRVSDITGLPT 83
OY 144 --SKNSANRELOKRLKLVKDNKKNRKPKRPVVDAGSGLDNGDFKVTPTDSTQ 201
DB 84 DPTSNCPN-----PYGNLVSRSCTGTIPN---TYAGYRSN-ETKTEPSANT-- 129
OY 202 HNKVSTSPKTTAKP-----INRPSLPNSDTSKR-----TSLIVN 238
DB 130 YAGYRSNENKTEPSANTNEFLVDPKINAPCSENSEFQOGIIFDMGKVIYPTKCAV 189
OY 239 KETTVKETTETNKQSTIDGKREKTSAKETOSIEKTSADLAPTSVLAKLPKAEETT 298
DB 190 KHTTTTTTTTTTTTTT-----TTTT-----TTT 211
OY 299 KGPALTPKPEPTTPREKPASTPKPEPTTKISAPTTKREKAPTTKSAPTTPKEPAPY 358
DB 212 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 271
OY 359 TTKREKAPTTKREKAPTTKREKAPTTKSAPTTPKREKAPTTKREKAPTTKREKAP 418
DB 272 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 331
OY 419 TPTTKREKAPTTKREKAPTTKREKAPTA PKKAPTTKREKAPTTKREKAPTTKREKAP 478
DB 332 TTTTTTKKPTTTTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTT 377
OY 479 KEKAPTTKSAPTTKKREKAPTTKSAPTTPKE----- 510
DB 378 TKPTTTTTTTTTTTTKKPTTTTTTTTTTTTSETSVIKPEMCWLEKNGCEAKGAYVYG 437
OY 511 -----PSPTT-----TKE 518
DB 438 VIGKGRLENGMAFTMINDTTHVFRKVKVGVNTISVRCRKAGAKLEPPDRSLDFTIP 497
OY 519 P-----APTTPKEAPTTPKKAPTTPKK----- 542
DB 498 PVAGHNSCIIIVSGDGKIHVSPGSKDVSLIS-----APIQPELENEVYCDCTAKYG 553
OY 543 ---PAPTTPKEAPTTTKKPAFTAPKREKAPTTPKETAFTPKKILPTTPEKLAFTPEKP 599
DB 554 AIHSGYQISADFTVTTTAKPTTT--TGAFQGPTTTTTGSPSKPTTTTTTKATTTT-- 607
OY 600 APTTPEELAPTTPEEPTPTTPEKAPTTPKAAAPNTPKREKAPTTKREKAPTTKREKAP 659

```

```

DB 608 --TTLNPTITTTTOKPTTTT-----TT-----KVGKRPATITTTTLKPIVTTTTKATTT 656
OY 660 PKETAPTTPKGTAPTTKREKAPTTPKKPAK-----ELAPTTKEPS----- 702
DB 657 TTTTPTT-----TTTTKRDGMTTTTTPLPDIGDIEITPIEKMDKTYRMIVDSGL 711
OY 703 --TSDKAPTTPKG--TAPTTKREKAPTTKREKAPTTKREKAPTTKREKAPTTKREKAP 756
DB 712 LIDSNDDEIPSSQAGIADTSNLFVOTHKSTGLPIDPMGLPDDKSGMLVHPINQNM 771
OY 757 APKELAPTTKGPSTUSDKAPTTPKKAPTTKREKAPTTKREKAPTTKREKAPTTKREKAP 816
DB 772 SGLSVYLAANKNLVDTDETGY--LPIDTGLVGLDVSILPFN--PETGELFDPIDSDI 827
OY 817 SPT-----TKETPTTKS-----PDEST- 836
DB 828 MNGTTAGIVSGISASESILSOKSALIDPATNMVYGERGLINPATGYMIGFLGSPQTO 887
OY 837 -----PELSAPPTPKALENSPKPEGVPTTKTPAATKPEMTTAKDKTTER--- 881
DB 888 FSEPEIEDGIIPEVAANAKFKLSIP--PSVP-----ESIDE-----KDKIDISE 934
OY 882 -----DLRTTPTTTAPKKTETATTTKTESKITAT--TQVSTTTQD 926
DB 935 LMVDIESGRILGOVSKRPISAGLDLNPIMKTPQDTSVGKPLDPTTGLPFPNPTGHL 994
OY 927 TTFKTTITLKTTLAPKVTTKKTYTTEIMKP--EETAKRD-----RATNSKATT 977
DB 995 INPTNNMTDSSPAGAKYKKAVANSGIKTDNVGLPVGELTGLPKDGSIDIPNSTTGELVD 1054
OY 978 PKPOKPLAKRKPTSTKPKTKMP-----RVKPKTTPPKRMTS 1016
DB 1055 PSTGKPIINNSTAGIVSGKGLPPIEDENGLFDPSTNLPIDGNNOVNPENINSYSGSTS 1114
OY 1017 --TMPE-----INPISRIAEAMLQTTTPRNOTPNSKL 1046
DB 1115 GTTKPKPGITVNGGAVVPDEAKDAQDKGKDLIVPTPMSINKDPVTNTOVSNTIGNT-- 1172
OY 1047 VEVPKSESDAGAEGETPHMLLRPHVMEVPPD-----MDLVRPNNGIILNP-- 1096
DB 1173 --INP--ETGKVIPLGSLPGLNPSFNTPOQDEITGKPKVDVTYGLPYDPSGELIDPAT 1228
OY 1097 -----MISDETINLNGK-----PVDDLTLRLNGTLV 1122
DB 1229 KLPIPSVAGDEILTEVLNITTTDEVGLPRLDLETGLPBDPVSGLPQLPNCGLV 1281

```

```

RESULT 5
US-09-103-429A-3
: Sequence 3, Application US/09103429A
: Patent No. 6187558
: GENERAL INFORMATION:
: APPLICANT: Granados, Robert R
: APPLICANT: Wang, Ping
: TITLE OF INVENTION: A NO. 6187558el Invertebrate Intestinal Mucin
: TITLE OF INVENTION: CDNA and Related Products and Methods
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Brown, plinisi & Michaels, P.C.
: STREET: 118 No. 6187558th Tioga
: CITY: Ithaca
: STATE: NY
: COUNTRY: USA
: ZIP: 14850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/103,429A
: FILING DATE: 24-JUN-1998

```


212 271


```

OY 437 TPKEPAPLAKKPA--PTTPKEBA--PTTPKEBAPTTPKEPSPTTP--KEPAPTTPK 487
Db 197 YP-----PTYAKKSYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYSLIKAKSPYPTPK 252
OY 488 SAPPTTKEPA-----PTTTSAPPTTPKEPSPTTTPKEBA--PTTPKEBA-----PTTPKK 534
Db 253 AKPTTKAKPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPK 308
OY 535 PA--PTTPKEBA--PTTPKEBAPTTPKPAAPLAKKEBAPTTPKEBAPTTPKILPTTPPK 590
Db 309 PSYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPK 362
OY 591 LAPPTPE--KPAPPTPEELAPTTPKEPT--PTTPKEBA--PTTPKAAAPNTPKKEBAPTTPK 645
Db 363 AKPTPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPK--AKSP--PTYKA 413
OY 646 EPA--PTT-----PKPAPPTTPKEBAPTTP--PKGAPPTTPKEBAPTTPKPAKELAPTTPK 698
Db 414 KPTPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPK 473
OY 699 EPTSTSDKPAPTTPKPAAPLAKKEBAPTTPKEBAPTTPKGTAPLTPKEBA--PTTPKKP 756
Db 474 KPTPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPK--YPTPYKAKSPYPTPYKAK 521
OY 757 AKELAPTTPKGTSTSDKPAPTTPKETAAPTTPKEBA--PTTPKKPA--PTT-----PETP 809
Db 522 TYK--AKPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPK--PTYKAKSPYPTPYKAKSP 576
OY 810 PPTTSEVSTPTTPKEPTTPHKSPTDESTPELSAEPPTPKALEMSPKPGPTTKTAAATKE 869
Db 577 PTYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPYKAKSPYPTPK-----KPTYPST--YKAKPS 620
OY 870 MTTTAKDTERDIRTTPETTAAP 894
Db 621 YPTTHAKPTPYKAKSPYPTTHAKAP 645

```

```

RESULT 10
US-09-219-849-5
; Sequence 5, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBRORCK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELIE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; OTHER INFORMATION: amino acid sequence
US-09-219-849-5

```

```

Query Match 6.2%; Score 452; DB 4; Length 960;
Best Local Similarity 25.6%; Pred. No. 3.3e-21;
Matches 173; Conservative 60; Mismatches 309; Indels 134; Gaps 41;

```

```

OY 290 PTPEAETTTGAPLTTKEPTTPPTTPKEBASTTPKEPTTIIKSAPTTPKEBAP--TTTKS 347
Db 72 PGAPGPGSRDGP--PGAPGAPG--PGSRDGPAGG--PAGPGRSDPGPAPGPA 126
OY 348 APPTPKBAPTTPKEBA--PTTPKEBAPTTPKEBAPTTPKSAPTTPKEBAPTTP--KKA 403
Db 127 GPGRSDPGPAGPAGPAGPGRSDPGPAGPAGPGRSDPGPAGPAGPAGPAGPAGPAGP 186
OY 404 PTTPKEBAPTTP--KEPTTPTPKEBA--PTTPKEBAPTTPKEBA-----442
Db 187 P--PGAPGAPGPGSRDGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 244
OY 443 -----PTAPK--KPAPTTP--KEBAPTTPKEBA--PTTPKEPSPTTPKEBAPTTPKS 488
Db 245 PAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 298
OY 489 APPTTPKEBAPTTPKSAPTTPKEPSPTTPKEBAPTTPKEBAPTTP--KKPAPTTPKEBAP 545
Db 299 GPGRSDGP-----PGAPGAPGPGSRDGP--PGAPGAPGPGSRDGP--PGAPGP 348
OY 546 TTP-----KEBAPTTPKKA-----PTAPKEBAPTTPKEBAPTTPPKILPTTPKEBAPT 595
Db 349 AGPGRSDGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 404
OY 596 PEKPAPT--TPEELAPTTPKEPTTPPEEAPTTPKAA--PTTPKEBAPTTP--KEBAPT 650
Db 405 SRDGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 462
OY 651 TPKEBA-----PTTPKET--APTTPKGT-----APTTPKEBAPTTPKPA--APKELA 693
Db 463 GAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 522
OY 694 PTTPKEPTSTSDKPAPTTPKETA--PTTPKEBAPTTP--KEBAPTTPKETA--PTTPKEP 748
Db 523 PGSRDGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 580
OY 749 APPTPKKPA-----KELAPTTPKGTST--TSKPAAPTTPKETAAPTTPKEBAPTTPK 800
Db 581 GP--PGAPGAPGPGSRDGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 636
OY 801 PAPTPP-----ETTPPTTSEVSTPTTPKEPTTPHKSPTDESTPELSAEPPTPKALEN-----S 851
Db 637 PGAPGPGSRDGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 695
OY 852 PKPQVPTTPKPAATK 867
Db 696 SRDGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 711

```

```

RESULT 11
US-09-041-886-23
; Sequence 23, Application US/09041886
; Patent No. 625872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Shiroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886

```

```

FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,615
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1185 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-041-886-23

```

```

Query Match 5.8% Score 425.5; DB 4; Length 1185;
Best Local Similarity 21.8% Pred. No. 2,1e-19;
Matches 240; Conservative 119; Mismatches 437; Indels 303; Gaps 52;

```

```

QY 85 SCITKSTTRSPKPPKPKKKKVKVISEETIEHSVSENOESSSSSSSSSTIMKISS 144
DB 9 SMSMRGRKKEAPR-----EELRSGRASPGGYSTSSDG-----KAEKS 50
QY 145 KNSAARELOKKLVKNDKNTKPKPPVDEAGSGLDNGDFKYTPDTSTGHNK 204
DB 51 ROTA-----KKAVEASTPK--VNKQGRS-----EISESESE 82
QY 205 VSTSKRTTAKPINDPRLSPNSDSEKSLTVNKETVETKEITTNKQSTGKEKT 264
DB 83 ETNAKKTTEQELPRQSPDLSDGRSL--NDGSSDRIODQDRST-----P 133
QY 265 SAKENQSTSEKSAKLAPTSKVLAKPKPKATTGKALTPPKERTPTTP--KEPASTP 322
DB 134 STYSGSEVNDSS-----DSSSGLSQGR--RPYHAPPLPSPQPPSTP 176
QY 333 KEPRTTAKSAPTTKEPAPTTPKAPTTKEPAPTTPKAPTTKEPAPTTPKAP 382
DB 177 ROPEASF-----EPHPSVTPTGYHAPHEP--PTSMFOAP--PCAPRHPOLYPGGT 224
QY 383 --TTKSATTPK-----EPAPTTPKAPTTKEPAPTTPKAPTTKEPAPTTPK 424
DB 225 GGVLSGPRMGKGGGAASVGGPNGKQHPPTTPISVSSGASGAPTPR-----PTTPY 280
QY 425 EPAPTKEPAP-----TTKEPAPTA--PKKAPTTKEPAPTTPKAPTTKEPAPT 477
DB 281 GGGULPSAPRANPNHTPLPPRLRLPLNNAASAP--PGIGAQPLRGHL--PSPYA 334
QY 478 PKE-----PAPTTKSAPTTKEP--APTTPKSAPTTTP--KEPSTTPKEPAPTKEPA 528
DB 335 MGOQMGILPRGP--EKGPLLAPSHSLPRASSANAPARHFRFYSSSSSSSAASSSSS 391
QY 529 PTPPKKRAPTTKEPAPTTPKE--PAPTTKKRAPTAPEKEDAPTTKEPAPTTPK 584
DB 392 SSSSASFPAS--QALPSYPHSFPPTSL-----SVSNQPR--PKYTQPSLPQAVMSQ 440
QY 585 --PTTP--KLAPTTKEPAPTTPPEELAPTTPEERTP----- 618
DB 441 GPPPPPYGRLNANVHGRPPRSGASTAHPRVSTNNHHHQQQQQQQQQQQHHG 500
QY 619 --TPPEAP-----TPPKAAPN-----TPKEPAPTTPKEP-----APTPE 654
DB 501 NSGPPRPGAPRPLEGGSSHHNAPVMSPLSLRYPRGRAHLRPPHSQVSYGAGNG 560
QY 655 PAPTPEKAPTTPKSTAPTTKEPAPTTPKAPKELAP-----TTTKPTSTSKAPPT 711
DB 561 PAVSSSSSSSSSSSSGYS--HPSGSGPGGADYPPRPVPTVTSATLSTYATAVAS 618
QY 712 TPAGTAPTTKEPAPTTPKEPAPTTPKSTAPTTLKEPAPTTPKAPPELAPTTTKGPTS 771
DB 619 SPAGYKTAAPRPPYKGRAPSGAVKATATPPGYKRGSP-----PSRPTGTPPGYRG-- 670

```

```

QY 772 TTSKPAPTTPKETAAPTTPKEPAPTTPKAPKAPTTPEPTPTTSEVSTPTTKPTTHNS 831
DB 671 -TSPRAGCTKPKSPVY--GGGPLPAGPS--GLPSLPAPPAPASGPLS--ATQIKGE 724
QY 832 PDE--STPELSAEPPTPALENS--PKEPGYPTTKPATATKPEMTTAKOKTTERDRTTP 887
DB 725 PAEEVETPE--SPVPARSPSPKVVDPVPSHNASOSARKNKLDGPFNSCARSDLEYVP 781
QY 888 ETTAARKMTKETATTEKT--TESKITATTTQVSTTTTQDTTPFKITTTTLKTTLAPKVT 946
DB 782 --LEGSKLAKKRDVLVEKVRREAR-- 807
QY 947 TKTTITTTIMNKEETAKPKDRATNSKATTPPKPKAPKPTSTKPKTPRYKPK 1006
DB 808 -----EKEKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREK 1066
QY 1007 TTPPKMTSTWELNPTSLAEAMLOTTTPRNPQNSKLVENPKSEDPAGAGGETPHM 1066
DB 849 LGVPYPR-----PPEPPGSAVA-----TVPPYLGPDTALRT--LSEYA----- 885
QY 1067 LRPVFMPEVTPDMYLP 1085
DB 886 --RPHVSPGNRHPYYP 902

```

```

RESULT 12
US-08-479-537A-5
Sequence 5, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHI, Richard
APPLICANT: HAREVENI, Maria
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,537A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 90/13101
FILING DATE: 23-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021

```



```

APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note= "The amino acids spanning
128 to 1727 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
FEATURE: repeats varies from 1 to 40."
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 - Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CGA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y - Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACG
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 - Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CGA
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-5

```

```

376 TKE--PAPTT-----KSAPTTKEPAPTTPK-----KPAPTTKEPAPTTKEPAPTT 422
181 APDXRPPGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVT 240
423 PRE--PAPTTKEP-----APTTKEPAPTPAK-----KPAPTTKEPAPTTKEPAPTT 469
241 APDXRPPGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVT 300
470 TREPPTTKEPAPTT--TKSAPTTTKEPAPTT-----TKSAPTTKEPAPTTTKEPAPTT 518
301 APDXRP--XPGSTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVT 359
519 PAPTTKEPAPTTPK-----KPAPTTKEPAPTTPK-----EPAPTTTKEPAPTTAPK 570
360 SAPDXRPPGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVT 419
571 TPKEPAP-----TTPKTLPTTPEKLAPTTPEKPAPTTPEELAPTTPEEPPTTPEEPAP 626
420 SAPDXRPPGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVT 471
627 TPKA-----AAPNTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTTPEKPAPTT 681
472 APXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSG 530
682 TTP-----KKPAKELAPTTTKEPSTSDKRPAP--TTPKGTAPTTTKEPAPTT 729
531 TAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSG 585
730 KEPAPTTPKG-----TAPTTTKEPAPTT-----KKPAKELAPTTTKEPAPTT 775
586 PPGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADK 644
776 KPAP--TTPK-----ETAPTTPEKPAPTT-----KKPAPTTPEPPTTSEVTP 819
645 RPPGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAAD 702
820 TTTKEPT-----IHKSPDESPELSAEPPTKALENSPEKPGVPTTKTPAA----- 865
703 DKRPAPGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSA 762
866 -TKPEPTTAKDK-----TTPRDLRTTPEPTTAPKKTATATTETTTSTKTATTTTOVTS 921
763 DKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSA 820
922 TTTODTTPKTTTTLTTLTTLAPKVTTKTKTTTTEIMNPEETAKPKDRTSKATTPRPQ 981
821 APDXRPPGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVT 868
982 KPTKAPKKTSTKTKPKTPRVKPKTTTTPPKKMTSTPELNPSTRIAEAMLQTTTRPQ 1041
869 -GSTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPX 911
1042 -PNSKLVENPKSEDDAGAEETPHMLLRPHVPEVTPMDVDYLPVYNOGIINPLMSD 1100
912 APXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSG 957
1101 ETNLCNGKRPVQGLTTP--LRNGLTVAFRGHYFMMLSPESPSPARRITTEWCIIPSPID 1155
958 VTSAPDXRPPGSGTAPXAHGVTSAADKRPXGSGTAPXAHGVTSAADKRPXGSGTAPXAH 1003

```

```

RESULT 14
US-08-479-537A-2
Sequence 2, Application US/08479537A
Patent No. 5861381
GENERAL INFORMATION:
APPLICANT: CHAMBOU, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREVEUENI, Mera
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

```

NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
 STREET: P.O. Box 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/479,537A
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: FR 90/13101
 FILING DATE: 23-OCT-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/FR91/00835
 FILING DATE: 23-OCT-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/039,320
 FILING DATE: 04-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/403,576
 FILING DATE: 14-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Teskin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 017753-025
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2035 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 128..1899
 OTHER INFORMATION: /note="The amino acids spanning 128 to 1899 constitute a repeated region wherein the repeat 20 amino acids, 17 of which are fixed. The number of such repeats varies from 1 to 40."
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 134
 OTHER INFORMATION: /note="Amino acid 134 is X1 = Xaa
 OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 144
 OTHER INFORMATION: /note="Amino acid 144 is Y = Xaa
 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC, ACA, or ACG; and Asn = AAT or AAC."
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 147
 OTHER INFORMATION: /note="Amino acid 147 is X2 = Xaa
 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
 OTHER INFORMATION:
 FEATURE:
 NAME/KEY: Peptide
 LOCATION: 1..21
 OTHER INFORMATION: /note="Amino acids 1 to 21 are a
 OTHER INFORMATION: 21 amino acid precursor sequence."

US-08-479-537A-2
 Query Match 5.8%; Score 424.5; DB 2; Length 2035;
 Best Local Similarity 23.7%; Pred. No. 4.6e-19;
 Matches 253; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

Qy	233	TSLTVNKEETVETKEETTTNKOVTSDCKEKTTSKKEQSLIEKTSAND-LAPTSKVIAPKT	291
Db	16	TVLVV-----VTCGSHASSTPGKEKETSATQSSVSSSTENKAVSMTSVLSHS	65
Qy	292	P-KAETTTKGP--ALTTPKEP-----TPTPKEPAPTTPKEPTTTKSA	334
Db	66	PGSSSTTGQODVTLAPATERPAGSAAATWQODVTSVPTRALGSTTPPAHDVT---SAP	122
Qy	335	TPPKKEPAPTTT-----KSAPTTTPKEP-----APTTPKEPAPTTPKEPAPT	375
Db	123	--DNKPAPGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVT	180
Qy	376	TKE--PAPTTT-----KSAPTTTPKEPAPTTPK-----KRAPTTTPKEPAPTTPKEPTTT	422
Db	181	APDXRPXPGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVT	240
Qy	423	PKE--PAPTTKEP-----APTTPKEPAPTAPK-----KRAPTTTPKEPAPTTPKEPAPT	469
Db	241	APDXRPXPGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVT	300
Qy	470	TKPEPTTPKEPAPT--TKSAPTTTPKEPAPT-----TKSAPTTTPKEPAPTTPK---E	518
Db	301	APDXRP--XPGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXGSTAPXAHGVT	359
Qy	519	PAPTTTPKEPAPTTPK---KRAPTTTPKEPAPTTPK---EPAPTTTPKRAPAPKEPAPT	570
Db	360	SAPDXRPXPGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXPGSTAPXAHGVT	419
Qy	571	TPKETAP-----TPPKLTTPTPPEKLAPTTPKEPAPTTPPELAPTTPEEPTTPPEPAPT	626
Db	420	SAPDXRPXPGSTAPXAHGVT---APDXRPXGSTAPXAHGVTSAADKRPX---PGST	471
Qy	627	TPKA---AAPNTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT	681
Db	472	APXAHGVTSAADKRPXGSTAPXAHGVTSAADKRP--XPGSTAPXAHGVTSAADKRPXG	530
Qy	682	TPP-----KKRAPKELAPTTTPKEPAPTTPK-----KRAPKELAPTTTPKEPAPTTP	729
Db	531	TAPXAHGVTSAADKRPXPGSTAP--XAHGVTSAADKRPXGSTAPXAHGVT---APDXR	585
Qy	730	KEPAPTTPKPG---TAPTTTPKEPAPTTP-----KKRAPKELAPTTTPKPGTSTSD	775
Db	586	XPXGSTAPXAHGVTSAADKRPXPGSTAPXAHGVTSAADKRPXPGSTAP--XAHGVTAPDX	644
Qy	776	KRAP--TTPK---ETAPTTTPKEPAPTTP-----KKRAPTPPEPTTPPTSEVSTP	819
Db	645	RPXPGSTAPXAHGVTSAADKRPXGSTAPXAHGVTSAADKRPXG--TAPXAHGVTSA	702
Qy	820	TTTKPEPTT-----IHSPESTPELSAETPRALENSKEPDEVPTTKTPAA-----	865
Db	703	DXRPXGSTAPXAHGVTSAADKRPXPGSTAPXAHGVTSAADKRPXPGSTAPXAHGVTSA	762
Qy	922	TTTQDTPPKITLTKTTTLAPVTTTKTITTTTEIMNKPEETAPKDRATNSKATTPKQ	981
Db	821	APDXRPXPG---STAPXAHGVT-----SAPDXRPXGSTAPXAHGVTSAADKRPX	868
Qy	982	KPTKAPKRPSTYKKRTPMRVAKKPTTTPPKKSTMPELPTRIAEAMQTTRNQT	1041
Db	869	-GSTAPXAHGVTSAADKRP--XPGSTAPXAHGVTSA--PDKRPX-----PGST	911
Qy	1042	-PNSKLVENPNSSEDAGGEGTTPMLLRPHVFMEVTPDDMDYLRPVNQGIIINPMLSD	1100

DB 912 APXAGVTSAPDXRPGSTAPXAH-----GVTSAPDXRPGSTA---PXAHG 957
QY 1101 ETNIGCKRPVDGLTT-LRNGTLVAFRGHYFWNLSPFSPSPARITTEWGIPEID 1155
DB 958 VTSAPDXRPGSTAPXAHGVTSA-----PDXRPGSTAPXAHGVTSA 1003

RESULT 15
US-09-083-116-2
Sequence 2, Application US/09083116
Patent No. 6203795
GENERAL INFORMATION:
APPLICANT: CHAMRON, Pierre
APPLICANT: KIENY, Marie-Paule
APPLICANT: LATHE, Richard
APPLICANT: HAREJVENT, Maira
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083.116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/479,537
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR91/00835
FILING DATE: 23-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/039,320
FILING DATE: 04-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,576
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 01753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1899
OTHER INFORMATION: /note= "The amino acids spanning
OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
FEATURE: repeats varies from 1 to 40."
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note= "Amino acid 134 is X1 - Xaa
OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CCT,

OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note= "Amino acid 144 is Y - Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note= "Amino acid 147 is X2 - Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-09-083-116-2

Query Match 5.8%; Score 424.5; DB 4; Length 2035;
Best Local Similarity 23.7%; Pred. No. 4.6e-19;
Matches 255; Conservative 69; Mismatches 511; Indels 241; Gaps 53;

QY 233 TSLVNNKTEVERKETTTKQSTDEKERTSAKETOSIEKTSARD-LAPTSKVLAKPT 291
DB 16 TVLTV-----VSGSHASTPGEKETATGORSVSPSTEKNASMSVSSSHS 65
QY 292 P-KAETTTKGP-ALTPKRP-----TPTPKREASTTPKEPTTTKSA 334
DB 66 PGSSSTTGQDVTLANATPRASGSAATWGDVTSVTPRALGTTTPAHDT---SAP 122
QY 335 TTPKEPAPTT-----KSAPTTKRP-----APTTPKEPAPTTKREPAPT 375
DB 123 -DNKRAPGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVT 180
QY 376 TKE-PAPTT-----KSAPTTKREPAPTTPK-----KRAPTTKREPAPTTPK 422
DB 181 APDXRXPGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVT 240
QY 423 PKE-PAPTTKEP-----APTTPKEPAPTAPK-----KRAPTTKREPAPTTPK 469
DB 241 APDXRXPGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVT 300
QY 470 TKEPSTTPKEPAPTTPK-----KSAPTTTKEPAPT-----TKSAPTTKREPSTTPK 518
DB 301 APDXRXPGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVT 359
QY 519 PAPTTKREPAPTTPK-----KRAPTTKREPAPTTPK-----EPAPTTPKRAPTAKEPAPT 570
DB 360 SAPDXRXPGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVT 419
QY 571 TKEETAP-----TPPKKLPTTPKEAPTTPEELAPTTPEEPPTPEEPAPT 626
DB 420 SAPDXRXPGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVT 681
QY 627 TPKA-----AAPNTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 681
DB 472 APXAGVTSAPDXRXPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGS 530
QY 682 TTP-----KKRAPKELAPTTTKEPTSTSDKRAP-----TPPKGTAPTTKREPAPTTP 729
DB 531 TAPXAGVTSAPDXRXPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGS 530
QY 730 KEAPATTTPK-----TAPTTTKEPAPTTP-----KKRAPKELAPTTTKEPTSTSD 775
DB 586 XPXGSTAPXAGVTSAPDXRXPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOX 644
QY 776 KRAP-----ETAPTTKREPAPTTP-----KKRAPTTPEPTTPPTSEVTP 819
DB 645 RXPXGSTAPXAHGVTSAPOXRPXGSTAPXAHGVTSAPOXRPXGS--TAPXAHGVTSAPO 702

```

QY 820 TTTKEPTT-----IHKSPDESTPELSAEPYPPKALENSPKPEGVPTTKTPAA----- 865
Db 703 DXRPXPGSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA PDXRPXPGSTAPXAHGVTSA PD 762
QY 866 -TKPEMTTAKDK---TTEBDLRTTPETTAARPKMTKETATTEKTESKITATTOVTS 921
Db 763 DXRPXPGSTAPXAHGVTSA PDXRPXPGST-APXAHGVTSA PDXRPXPGSTAPXAHGVT 820
QY 922 TTTODTTPFKITTLKTTTLAPKYTTTKITTTTEIMNKPEETAKPKDRATNSKATTPK 981
Db 821 APDXRPXPG-----STAPXAHGVT-----SAPDXRPXPGSTAPXAHGVTSA PDXRPX 868
QY 982 KPTAPKPKPTSTKPKPTMPYRKPKTPTPRKMTSTMPDELNPTSRIAEAMLQTTTRPNOT 1041
Db 869 -GSTAPXAHGVTSA PDXRP--XPGSTAPXAHGVTSA-PDXRPX-----PGST 911
QY 1042 -PNSKIVEVNPKSEDAGAEGETPHMLLRPHVEMPEVTPMDYLPRVNOGIIINPMLSD 1100
Db 912 APXAHGVTSA PDXRPXPGSTAPXAH-----GVTSA PDXRPXPGSTA---PXAHG 957
QY 1101 ETNIONGRPV DGLTT-LRNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSPID 1155
Db 958 VTSAPDXRPXPGSTAPXAHGVTSA-----PDXRPXPGSTAPXAHGVTSA PD 1003

```

Search completed: April 26, 2002, 16:25:33
 Job time: 473 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2002, 16:31:27 ; Search time 114.61 Seconds

(without alignments)
844.095 Million cell updates/sec

Title: A43
Perfect score: 6814
Sequence: 1 MAWKTLPTLYLLLSFVYIQ.....ARATITRSGQTLKRWYVNC 1270

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR_68:*
2: PIR1:*
3: PIR2:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1065	15.6	3020	2 A43932	mucin 2 precursor,
2	950	13.9	1664	2 T18262	S-layer protein -
3	855	12.5	1489	2 T31108	cyst germination s
4	802	11.8	1274	2 T16251	hypothetical prote
5	798.5	11.7	1367	1 S48478	glucan 1,4-alpha-g
6	784	11.5	1188	2 S49915	extensin-like prot
7	776.5	11.4	2187	2 T30826	nascent polypeptid
8	672	9.9	1344	1 A35175	mucin 1 precursor,
9	671	9.8	1151	2 T18535	high molecular mas
10	670	9.8	1229	2 T25697	hypothetical prote
11	651.5	9.6	3570	2 T45025	mucin MUC5B, trach
12	641.5	9.6	3507	2 T34513	hypothetical prote
13	635.5	9.3	7962	2 T18346	elastin titin - hu
14	633	9.3	489	2 T11622	extensin class 1 p
15	632	9.3	761	2 C84672	hypothetical prote
16	630.5	9.2	990	2 T51618	nucleolar phospho
17	626.5	9.2	971	2 T19431	hypothetical prote
18	625	9.2	6642	2 T29757	protein UNC-89 - C
19	622.5	9.1	839	2 T29757	hypothetical prote
20	608	8.9	801	2 T29018	hypothetical prote
21	607.5	8.9	924	2 S27923	gene lrp3 protein -
22	605	8.9	379	2 S50125	larval glue protei
23	559.5	8.2	350	2 S22456	hydroxyproline-ric
24	556.5	8.2	856	2 T16543	polyphenolic adhes
25	555.5	8.2	875	2 S23760	hypothetical prote
26	555.5	8.2	1630	2 A53577	ascites stialoglyco
27	554	8.1	2232	2 T34434	hypothetical prote
28	551	8.1	620	2 S06733	hydroxyproline-ric
29	545	8.0	873	2 A47283	calphostin - fruit

30	542	8.0	369	2 S20500	hydroxyproline-ric
31	540.5	7.9	756	2 T27642	hypothetical prote
32	538.5	7.9	416	2 J00465	extensin precursor
33	533	7.8	1087	1 QFMSH	neurofilament trip
34	530.5	7.8	1162	2 JH0557	exo-alpha-stallidas
35	530	7.8	865	2 A47282	calcium-binding pr
36	522	7.7	328	2 J00985	hydroxyproline-ric
37	522	7.7	1459	2 T32271	hypothetical prote
38	518	7.6	813	2 S70795	vasa protein precu
39	518	7.6	866	2 T35462	membrane glycoprot
40	518	7.6	1072	1 A37221	neurofilament trip
41	512.5	7.5	867	2 T45463	membrane glycoprot
42	509.5	7.5	662	2 A45155	mucin FTM-C.1 - At
43	508	7.5	1832	2 T31113	mucin-like glycopr
44	505	7.4	700	2 A54641	interspersed repea
45	504.5	7.4	606	2 A43427	neurofilament trip

ALIGNMENTS

RESULT 1
A43932
mucin 2 precursor, intestinal - human (fragments)
N:Alternate names: mucin SMUC-41
C:Species: Homo sapiens (man)
C>Date: 10-Mar-1993 #sequence, revision 12-Apr-1996 #text change 05-Nov-1999
C:Accession: A49963; #sequence, revision 12-Apr-1996 #text change 05-Nov-1999
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A>Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t
A:Reference number: A49963; MUID:94132002
A:Accession: A49963
A:Molecule type: mRNA
A:Residues: 1-639 <GU1>
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A>Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up
A:Reference number: A45106; MUID:93016075
A:Accession: A45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 626-1895 <GU2>
A:Cross-references: GB:M94131; NID:9186395; PIDN:AAA59163.1; PID:9186396
A>Note: sequence extracted from NCBI backbone (NCBIP:116706)
A:Accession: B45106
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 2037-3020 <GU3>
A:Cross-references: GB:M94132; NID:9186397; PIDN:AAA59164.1; PID:9186398
A:Experimental source: colon
A>Note: sequence extracted from NCBI backbone (NCBIP:116698)
R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A>Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A:Reference number: A43932; MUID:91358717
A:Accession: A43932
A:Molecule type: DNA
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
A:Cross-references: GB:M74027; NID:9188863; PIDN:AAA59875.1; PID:9188864
A>Note: sequence inconsistent with the nucleotide translation
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A>Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl
A:Reference number: A33532; MUID:89197956
A:Accession: B33532
A:Molecule type: mRNA
A:Residues: 1916-2193 <GU4>
A:Cross-references: GB:M22405; NID:9188873; PIDN:AAA36334.1; PID:9188874
A:Experimental source: intestine
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

J. Clin. Invest 87, 77-82, 1991
A:Title: Human bronchus and intestine express the same mucin gene.
A:Reference number: A61257; MUID:91086481
A:Accession: A61257
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
A:Experimental source: bronchus
A:Experimental source: intestine
A:Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-terminus
A:Reference number: P00328; MUID:92198477
A:Accession: P00328
A:Molecule type: mRNA
A:Residues: 2328-2468 <XUG>
A:Cross-references: GB:M86523
A:Experimental source: small intestine
A:Accession: P00329
A:Molecule type: protein
A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
C:Genetics:
A:Gene: GDB:MUC2
A:Cross-references: GDB:120203; OMTM:158370
A:Map position: 11p15.5-11p15.5
C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von Willebrand factor type C repeat homology
C:Keywords: glycoprotein; intestine; tandem repeat
E:2766-2634/Domin: von Willebrand factor type C repeat homology <VWC>

Query Match	15.6%;	Score 1065;	DB 2;	length 3020;
Best Local Similarity	28.7%;	Pred. No. 2e39;		
Matches 370;	Conservative	79;	Mismatches 464;	Indels 376;
				Gaps 49;

[illegible][illegible]

Query Match	13.9%	Score 950;	DB 2;	Length 1664;
Best Local Similarity	31.0%	Pred. No. 1.2e-34;		
Matches 312;	Conservative 104;	Mismatches 361;	Indels 230;	Gaps 55;
OY 193 VLAKPFP-KAETTKCPALTTKEPPTTKPEKASTTKPEPTTKKSAFTTKPEAPPT 251				
: : :	:	:	:	:

Db 758 VVIOAPAKASDEPIPTDPSDEPTPS-----DEPTPS---DEPTPSDEPTPSD 804
Qy 252 TKSAPPTPKAPPTTKEPAPTTPKEPA-----PTTKEPAPTTPKSAPTTPKEPAPTTPK 307
Db 805 EPTPESEPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 307
Qy 308 KPAETTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 308
Db 865 SDEPTPSDEPTPS---DEPTPS---DEPTPS---DEPTPSDEPTPSDEPTPSDEPTPS 308
Qy 367 TTPKEPA---PTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 367
Db 920 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 367
Qy 420 -----PTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 367
Db 980 PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 367
Qy 475 APT---TPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 367
Db 1040 TPDSEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 367
Qy 528 EEPAPTTPKAAPNTPKAPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 367
Db 1095 EEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 367
Qy 585 EEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 367
Db 1155 EEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 367
Qy 644 PKGAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 367
Db 1209 -----DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 367
Qy 700 PAPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 367
Db 1254 PTPS---DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 367
Qy 759 PREPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 367
Db 1306 SDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 367
Qy 817 SKITATTQVSTTTQDTTPKITTLLKTTTLLKTTTLLKTTTLLKTTTLLKTTTLLKTTTLLKTTTLL 367
Db 1362 DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 367
Qy 877 ATNSKAT-----TPKPKAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 367
Db 1397 SGGSGSGSGGG 367
Qy 922 -----TSPMELNPTSRSLAEAMLQTTTRNQTNRNSKLVEVNPKSESDAGAGGEPHNM- 367
Db 1454 YLRGYPDGSGFRERNITRAEAAVIF-----AKLL-----GADESAGASASASYSD 367
Qy 974 LLRPH-----VEMPE-----VTPMDVILPRVNOGI----- 367
Db 1499 LAIDHMAAMAKPATSGKRGKRGKRGKRGKRGKRGKRGKRGKRGKRGKRGKRGKRGKRGKRGKRGK 367
Qy 1000 ---LIPMLSDETNICNG---KPVGLTLL-----RNGTLVAFRGHFWMLSPSPSPS 367
Db 1559 TIDISNPKFPD---CVGHAQGEIEKLTSLGYSIGPDT-----FKPON 367
Qy 1047 PARHTEVWGISPIDVIFRCNCEGKTFFFKD---SOYMFETDIND 1091
Db 1601 YIKRSEV---ALINALERGLNAPKLFPPVNESYNAF-GDIMD 1642

RESULT 3
T31108
Cyst germination specific acidic repeat protein precursor - Phytophthora infestans
C:Species: Phytophthora infestans (potato late blight agent)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31108

R:Goernhardt, B.
submitted to the EMBL Data Library, April 1998
A:Reference number: 220986
A:Accession: T31108
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1489 <GOE>
A:Cross-references: EMBL:AF061185; NID:g3851513; PID:g3851514; PID:AC72308.1
C:Genetics:
A:Gene: car90

Query Match 12.5%; Score 855; DB 2; Length 1489;
Best Local Similarity 31.8%; Pred. No. 1,6e-10;
Matches 362; Conservative 52; Mismatches 496; Indels 228; Gaps 51;

Qy 37 FERGREGCDACQCKYDKCCDYSFCAEVKN---KKNRTKKKPKPPVNDAGSGLD 93
Db 301 YEPSEDEAPTEGTY---VREETTAAPSDEDTYVAREVYVAPTEKPYDVEETTYVTE 357
Qy 94 NGDKVYTPDITST---OHNVSTSPKITTAKPINRPSLP-----P 132
Db 358 ESTYAPTKSETNAPTEKMHVAHIEKPCDTEVYMAPTEETTYAPTEETTYAPTEETTYAP 417
Qy 133 NSDT-----SKETSIVKKEETVETKETT---TNKQSTDGKSKTTSKETSQISAKTSAK 185
Db 418 TEETPYEETETTYTTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 477
Qy 186 DLAPTSKVLAKPKPAKAPTEKTPKAPALTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 245
Db 478 TYAPTEETTYAP---DETYASTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 533
Qy 246 EPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 296
Db 534 TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 592
Qy 297 ---TEKAPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 335
Db 593 PDETYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 652
Qy 336 TTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 387
Db 653 TYAPAEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 712
Qy 388 P-----APT---TKKAPT-----TKKAPT-----APT---TKKAPT---TPKPSPT 421
Db 713 PDETYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 772
Qy 422 TKKEP-----APTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 473
Db 773 TYEPTGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 832
Qy 474 PAPTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 530
Db 833 PTEET---TYTPEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 590
Qy 531 ---APT-----TPKAAAPNTPKAPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKE 569
Db 891 KETTYAPTEETTYASTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 950
Qy 570 TAPTPKGTAPT---TKKAPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 624
Db 951 YAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1010
Qy 625 PTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 660
Db 1011 EETTYAPAEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1070
Qy 661 KKP-----APKE---LAPT---TKGPTSTSDKAPAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 709
Db 1071 YEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP 1130
Qy 710 PTPPEPTTSEVSTPTTKAPT---TIHKSDESTPELSAETTPKA---LENSPKKEPG--- 763

```

Db 1131 EETTYAPTEETMYAPIEETTYGPTTEETTY-APPEATTYAPTEETTPAPTEETTYETCT 1189
Oy 764 --VPTTTPATAPTEEMTTTAKKOTTEEDLTTP-----ETTTAPKMTTKTATTTETKTES 817
Db 1190 TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYEPTTEETTYAP-----TEETTYEPTBE 1244
Oy 818 KITATTTQVSTTTODTT-----PPKITTT---LKTTLTAPVNTTKITTTTEIM 864
Db 1245 TTVAPEETTYAPTEETTYAPTEETMYAPIDEETTYGPTTEETTYAPPEATTYAPTEETPYA 1304
Oy 865 NKPEETAKPDRAK--NSKATTPPK--OKPTKAPKPKP-----STKRP-KTMR 908
Db 1305 PTEETTYEPTGTEETTYAPTEETTYAPTEETTYAPPEETTPYPAEESTSTVSTETKPCNTTEP 1364
Oy 909 VRKPTTPTPKMTSTMBELNPTSRIAEMLQTTTP--NOTNSKLVEYNPKSEDAK 964
Db 1365 TDEPTDEPTDE--PSDEPTDEPTDEPTDLPTEDEPTDCDNGINGIGVENKVRINNG 1420

```

```

RESULT 4
T16251
hypothetical protein F35A5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
C:Accession: T16251
R:Leinbach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: Z18485
A:Accession: T16251
A:Stetus: preliminary; translated from GB/EMBL/DDBU
A:Molecule type: DNA
A:Residues: 1-1274 <LEI>
A:Cross-references: EMBL:U46675; NID:q1166613; PID:q1166621; PIDN:AAB52641.1; GSPDB:GN00
A:Experimental source: Strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP:F35A5.1
A:Map position: X
A:introns: 1272/2

```

```

Query Match 11.8%; Score 802; DB 2; Length 1274;
Best Local Similarity 28.0%; Pred. No. 2.9e-28;
Matches 290; Conservative 102; Mismatches 414; Indels 230; Gaps 56;

```

```

Oy 71 KNRTKK-KP-----TP-----KPPVDEAGSGIDNGD-----FKVYTPDTSTQHMKVSTN 115
Db 274 KNPTKKMPWEDETPVEEVKEPPPEKAPVLAKKDPAPAKARDSPSKAAPKVEPS 333
Oy 116 PKTTAKPT-NPRPSLPNSDTSKETSILVYKETTVEKETTTTNKQSTGDEKKTTSK 174
Db 334 SPVVPPTPVKNPVKKKPPWEVDEFAEEVKKPSABEKTPVLKRRPEPSSSTTPSSDS 393
Oy 175 ETOSIEKTSADLAPTSVLAKPTPKAE-----TTTGP-----LTP 213
Db 394 PKKAPAVKRRDSSPKKATPLQADPKAOEVPPTPVKNPVKKYKPPWEVDEDEPVEEVKQ 453
Oy 214 KEPTPTTP-----KEPATITKEP-----TP--TTIKSAPTTKEPATITTKSAPTTPKEPA 264
Db 454 EAPKPKPPVLKRRKPAKDKATSKTSTPEPEKKDPVKRRDSSPKKAVAKPPSAQA-PA- 512
Oy 265 TTTKEPA-----PTTKEPATITTKEPAPT-----TTKSAPTTP 298
Db 513 TPVKNPVKKMPWEDETPADVDKPTDAKKTPTSLAKKPPAPAKESLAKRADTKAPAKP 572
Oy 299 KEP-----APTPKKAPATTPKE-----PAPTTPKEPTTPTPKKEPATITTKEPAPTTPK-EP 348
Db 573 RDPSPKVAAPAPAEKTPVLAKKKEPAGADSKTEPEKSPKRDSPKAVAPAKVPKTEV 632
Oy 349 APTAKKAP-----TTPKKEPATITTKEPAPTITTKEP-----SPTTKEPATITTKSAPT 398
Db 633 APAAVKKPEPISKPKDTAPKKAEPNSVVP-PTPVKNPVKKMPWEDEDDAPAPAPVSLPE 691

```

```

Oy 399 TTKKEPATITTKSAPTTP-KEPSPITTKKEPATITPK--EPAPTTPKKAP--TTPKKEPAT 453
Db 692 PEKK--TPVLAKKAPTTPDSEAAADPVSGPSKDPKLAKKAPVKKRDPSPKAVPIKAPK 750
Oy 454 TPKKEPATITTKKAPTA-PKKEPATITTKETAPTTPKKITPTTP-----EKLATTEKPA 507
Db 751 T--EVPAPVKKPEEVAKSPDSKPKAK-AEPNSP--VPPPTPVKNPVKKMPWEDEDDA 805
Oy 508 PTTPELATPTTEEP-----PTTPEEPATTPKKAAPNTPKKEPATITTKKEPATIT-PKE 561
Db 806 PAEPNVNPEPEKKTTPVLAKKTTPVKRDPSPKKAAPAKSTTDAPVSVKKPEVSKPE 865
Oy 562 PAPTTPKEPATITTPGNTAPTTLKEPATITPKKPAKELAPT--TKETPTSTSDKPAITTP 620
Db 866 PSKKAEPNSVVP--PTPVKNPVKKW-KPPWEDEDEPTEVKKKPE--PEKKTPLA 918
Oy 621 KGTAPTPKEPATITPKKEPATITPKGTAPTTLKEPATI-----TPKKAP-----KEL 666
Db 919 K-KEPEKPD-APKVAAPKRDSPSKAVE--KEPAKVAAPKRDLSPPKALIPANTOEA 974
Oy 669 APPTTKGPTS-----TTSDEP-----APTPEKET-----APTTPKEPATITPKKAP 711
Db 975 PPTPVKNPVKKMPWEDEDEPAEPVSAPEPEKKTTPVLAKKAPAKRDP--SPKKAAPV 1031
Oy 712 TPTTPPTTSEVSTPTTKETPTTIHKS-----DESTPELSA-EP--TPKALENSPKREG 763
Db 1032 AAK-PDPKITEV-PPTPVKNPVKKMPWEDEDESEPSAPEPEKKTTPVLAKKAPTPA 1089
Oy 764 V-----PTTKTPPAT-----KPEMTTAKKTERDURTPTT--TAAPK 802
Db 1090 TKPDSEAAADPVSGPTSPKPSKKAPEKPKTTDPRKDKLPSPAKKPEKAPPAAPK 1149
Oy 803 MKETATITTEKTESKITATTTQVSTTTODTTTPEKITTLKTTTLAPKYTTTKITITTE 862
Db 1150 KMKPVMDDEDEADPTVAPSKKPTDEPADPL-----GPKTKDPK----- 1193
Oy 863 IMNKPEETAKPKDRATNSKATTPKPKPTKAPKPKPTSTKPKTMPVVRKPK----- 913
Db 1194 -LNKKAPEKPTK-----PKPEVSKPEPKPTPEPKP-AAPKMKMPWEDEDEDE 1243
Oy 914 ---TPTPKMTSTMP 926
Db 1244 ADFTMPAKPKDPTDTP 1259

```

```

RESULT 5
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIK01
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C:Accession: S48478; A26877; B26877; S27281; Jc6123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48478
A:Molecule type: DNA
A:Residues: 1-1367 <ROW>
A:Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009.
A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
J. Bacteriol. 169, 2142-2149, 1987
A:Reference number: A91831; MUID:87194600
A:Accession: A26877
A:Molecule type: DNA
A:Residues: 1-242 <YAM>
A:Cross-references: EMBL:M16164; NID:q172522; PIDN:AAA35014.1; PID:g172525
A:Accession: B26877
A:Molecule type: DNA
A:Residues: 762-1331 <YAZ>
A:Cross-references: EMBL:M16165; NID:q172523; PIDN:AAA35015.1; PID:g172526
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Clares, M.G.; Jimenez, A.

```

FEBS Lett. 239, 179-184, 1988
A.Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Saccharomyces cerevisiae
A.Reference number: S27281; MUID:89031230
A.Accession: S27281
A.Molecule type: DNA
A.Residues: 1-31 <PAR>
A.Cross-references: EMBL:X13857; NID:g4551; PIDN:CA432069.1; PID:g4552
R.Lambrechts, M.G., Bauer, F.F., Marmur, J., Pectorius, I.S.
A.Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseudomyxomatous peritonitis in the mouse
A.Reference number: JG6123; MUID:96323237
A.Accession: JG6123
A.Status: nucleic acid sequence not shown
A.Molecule type: DNA
A.Residues: 1-1367 <LAMB>
A.Cross-references: GB:J03626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C.Genetics:
A.Gene: SCD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A.Cross-references: MIPS:YIR019C; SGD:S0001458
A.Map position: 9R
C.Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
F:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F:5-21/Domain: transmembrane #status predicted <TM1>
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match	11.7%;	Score 796.5;	DB 1;	Length 1367;
Best Local Similarity	27.9%;	Pred. No. 4.5e-28;		
Matches 313;	Conservative 108;	Mismatches 523;	Indels 177;	Gaps 50;

```

0Y      44 DCDADQCKKYYK---CCPDYESFCAEYK-DNKKRNKTKKKKPPKAPVYD- -AGSGLDNDF 97
Db      174 DLSGICNNYNOGHQSOTDPEGPGYWNIDCDNCGGTSSSTTSSSTSSSTTSSSTSSSTTSSSESSTT 233
0Y      98 KVTTPDSTTQHNKVSTSPKRTTAKPIND-----RPSLRPNDSOTKETSILVKNKETT 149
Db      234 TSSSTSSSTSSSTSSSTSSSTSSSTTAPAPPTTSCCKEKPRTPTTSCCKEKPRTPPHDDTT 293
0Y      150 VETRETTTNNKOSTGCKEKTSAKETOSIEKTSKDLAPTSYKLANPPEKAETT---TKG 207
Db      294 PCGRKTTTTSK-TCI---KKTTPVPPTS-SSTTESSAPV-----PTPSSSTTESSSA 342
0Y      208 PAITPKE-----PPTTPKPEASTTKEPRTTITKSAP---TPKPEAPPTTKAPT 257
Db      343 PVTSSSTTESSAPVPTPSSSTTESSSAPVSTTESSSAPVTSSTTESSSAPVPTPSSST 402
0Y      258 TPKEAPPTTKPEAPPTTKPEAPPTTKKAPPTTKKAPPTTKKAPPTTKPEAPPTTKPEB 317
Db      403 TESSSAPVTS---TTESSSAPVTS---STTESSAPVT---SSTTESSSAPVTSSTT 451
0Y      318 APTTPEP---TTPKPEAPPTTKPEAPPTTKKAPPAKKAPPTTKPEAP---TTPK 370
Db      452 ESSSAPVPTPSSSTTESSSAPVT---SSTTESSAP-VPTPSSSTTESSSAPVTSSTBS 507
0Y      371 EPAPPTTKEPSTTPKPEAPPTTKSAPPTTKKAP---TTPKSAPPTPEEPSTTPK 426
Db      508 SSAPVPTS-SSTTESSSAPAPPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESS 566
0Y      427 AP---TTPKPEAPPT-TPKKAPPTTKPEAPPTTKKAPPTTKKAPPAKPEAPPTPE 481
Db      567 TPTVSTTESSSAPVPTPSSSTTESSSAPVT---PSSSTTESSSAPAPPTPSSSTTESS 622
0Y      482 TAPTPPKLPTPEKLAPTTPEKPAPTPEELAPT-TPEEPPTTPEEPAD---TTP 535
Db      623 SAPVT---SSTTESSAP-VPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTE 677
0Y      536 KAAPPTPKPEAPPTTKPEAPPTTKPEAPPTTKETAPVT-TPKGAPPTTKKAPPTTKPKP 594
Db      678 SSAPVPT---SSTTESSSAPVT---SSTTESSSAPVPTPSSSTTESSSAPVPTPSSST 729
0Y      595 APKLAPTTKEPSTTSDKAPPTPKGAPPTTKPEAP---TTPKPEAPPT-TPKGAPT 650
Db      730 TESSSAPVPTS-SSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTE 788

```

```

QY 651 TLKRPAPTTKKRPAKELAPTTKGSTTSDKPAPTTKRETAPTTKREPAP-----TTP 705
Db 789 SSSAPVPTPSSSTTSSSVAPVPTPSSSSNITSSAPSSPSSSTRESSSVPTPSSSTTE 848
QY 706 KKRPAPTTPEPPTTSEVPT-----TTKEPTIHKSPDESTPELSAEPPTALENSPK 760
Db 849 SSSAPVSSSTTESVAPVPTPSSSNITSSAPSSI---PESSTTESFSTG-TVTPSSSK 904
QY 761 EBPVPTTKPAAITKPEMTTAAKDKTREDLTTPETTTAAAPKMKKEPATTEKTTESKIT 820
Db 905 YPESQTESVSSSTETTIVPTTKTTSVPTTITTTTVCSTGNSAGETTSGCCSKRYPT 964
QY 821 AT--TTPVSTTQDTPETPKITTLTKTLAPKVT---TKKTTTT-ELMKNKEETAKPK 874
Db 965 TIVPTTTTTSVTSSSTTTTTTVCSTGNSAGETTSGCCSKRYTTTTVPCSTPSETA--- 1022
QY 875 DRATNSKATTPKPKP-----TKAPKPKTKKPKTKMPPRVKPKRTPT 917
Db 1022 ----SESTTSPPTPVTVTVSVTVVTEYSTKPKGELITTFYVKNIPPTVLTTLAPT 1076
QY 918 PKKMTSTPELNPSTSRILAEALOTTTPPNOTPNSKLEVENPSEDAGAEGETPHMLRP 977
Db 1077 P--SVTTVNTPTT-ITTTVCSTGT-----NSAGETTSGCCSKPTVTTTVCPC 1120
QY 978 HVFMPEVTPDMQDYLPRVNPQIILNPMILDEFNICNGKRPVGDGTLTLNRGTLTIAFGHFW 1037
Db 1121 STGTGEYTTETATTLVTAVTTTVVTTSESSGTNSA-GKTTTGVTIKSVPTT-----YVT 1173
QY 1038 MTSPPSPAPARITLFWGJLPSPIDIVTRC-----NCEKRT 1074
Db 1174 TLASAPVTPPATN-----AVPTITTT--TECSATTAAGET 1207

```

RESULT 6
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465. NID:6600117. NID:010000000

Query Match	11.5%	Score 784;	DB 2;	Length 1188;
Best Local Similarity	29.9%;	Pred. N7;	1.7e-27;	
Matches 230;	Conservative	59;	Mismatches 359;	Indels 122;
				Gaps 33;
QY	78	PIPKPVPVDEAGSLDNGCFKVTTPDTSTGTHNNKYSTSKITTAAPINRBSL----	PPN	133
Db	460	PIPHSPPAD-----DYVPPTPVVECKSPSPATSPSOVOPPAASPTPVLVLSPPQ		510
QY	134	SDPSKSTSLTVAKNETVETKEETTNNKQSTDGKEIKTSAKEQSIERTKADLAPTQV		193
Db	511	APVGV-----SPPPVKTTSPAPRIG-----SPSPPPVSVV		541
QY	194	-----LAKETPKAETTTKGPALTTPKEKPTPTTKREPASTTKKEPPTTIKASAPTTPKPA		248
Db	542	SPPPVKSPPPPAVAVGSPPPPEKSPPPPAVAVASPPPVKSP--PPPTLVASPPPVVSP		599
QY	249	PTTTASAPTT--KEBAPTTKKEPAITTKKEBAPTTKKEBAPTTKKASPTTP--KEBAPT		305
Db	600	PPAPVASPPPVKSPSPPTPVASPPPPAVASPPPMKSPPTPVVSPSPPEKSPSPPP		659
QY	306	PKKPAPTPKKEBAPT-----PKE-----PTTTTKKEBAPTTKKEBAPTTKKEBAPTAKKP		356


```

Db 656 APSTAPPAHGVTSAPDTRPAGSTAP-PAHGVTSAPDTRPAPGSTAPPAHGVTSAPDTR 714
OY 698 KEPAPTP-----KKAPTPPETPTSEVSTPTTKEPT-----IHKSPD 740
Db 715 PAPSTAPPAHGVTSAPDTRPAGS--TAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPD 772
OY 741 ESTPELSAEPKALENSPREPGVPTTKPAA-----TKREMTTKAKD---TTEED 789
Db 773 TRPAPGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPD 832
OY 790 LRTTPEIT-----TAPKMTKETATTEKTESKITATTTQVSTTTQDTP--FKIT 840
Db 833 TRPAPGSTAPPAHGVTSAP--DTRPAPGSTAPPAHGVTSAPDTRPAGSTAPPAHGV 888
OY 841 TLKTTTLAPKVTYTKITTT--EIMNKPEETAKPKDRATNSKATTKPKPKAPKPKPT 898
Db 889 SAPDTRPAPGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRAP--GSTAPPAH 946
OY 899 STKKKTPPRKPKTKTTPTPPKMTSTPELNP---TSRIAEMLQT-TTRP--NOTPN 950
Db 947 VTSAPDTRP---APGSTAPPAHGVTSAPDTRPAGSTAPPAHGVTSAPDTRPAGSTAP 1003
OY 951 SKLVEVNRKSEDAEGETPHMLRPHVMEVTPDMDYLPVYVNGIILNPMLSDETIN 1010
Db 1004 AHGVTSAPDTRPAGSTAPPAH-----GVTSAPDNRPALGSTA---PVHNTVS 1049
OY 1011 ICKGRPVGLTLNGLTVAFRGHVFMMLSPSPPS 1046
Db 1050 ASGSASGSASTLVHNGTSARATTTTASSTPSPIS 1085

```

```

RESULT 9
T18535
high molecular mass nuclear antigen - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18535
R:Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick
A:Reference number: Z18955; MUID:9803440
A:Accession: T18535
A:Status: preliminary; translated from GB/EMBL/DDAJ
A:Molecule type: mRNA
A:Residues: 1-1151 <SH1>
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BA24137.1

```

```

Query Match          9.8%; Score 671; DB 2; Length 1151;
Best Local Similarity 25.9%; Pred. No. 1.5e-22;
Matches 284; Conservative 100; Mismatches 419; Indels 292; Gaps 55;

```

```

OY 79 TPKP-----PVVDEAGSLDNGDFKVTTPDTSTIOHKNVTS--PKITTAKPINRPS 129
Db 113 TPSPSGPAGTPPPSGAGAGAPKGDGTAOSGTRKSGADGPAADVYKATTA-ATAARP- 170
OY 130 LPPMSDSKESLSLVNKEVETVETKETTANK--QTSFGKEKTTSAKETOSIEKTSAKDL 187
Db 171 -----ASAASPTVPKATAATAVTAASOSAPKAAATDAATAA---SOSAPKATV-EV 219
OY 188 APTSKVLAKPPPKAETTKGPAITTPKEPTP--TPKPEASTTPREP-----TPITISAPT 242
Db 220 KPAANAAVAKKAAVTAATAAAKATAEAKPAVVTSPITIPCSAEAKPLTAASPTASKA--T 277
OY 243 TPKPEAPT-----TKSAPTPPKPEAPTTTK----- 268
Db 278 AEAKVVPATASIMATKVTAEAKPAFSPVKATITDKAVITAFKAGDVKPAVAOCAEA 337
OY 269 EPAPTTPKEPAPTTTKEPAPTTTKSAPT-----PKPEA-----PTTPKKAPPTTPK 315
Db 338 KRAPPPPPQOLPKAAAAAPFTGTELKPATAPPHGSPRANSHYVTPPNVPRAAATVP- 396
OY 316 EPAPPTPKPEPTTT-----PKPEAPTTTKEPAPTTTKEPAPTAAPKKAPPTTPKEPAPTTTKE 371

```

```

Db 397 -TAGAVPKASIGTTPAAPAPQOPV--KAAPVTPPSDOAVPRAATAAA---APVTPQ 448
OY 372 P---APTTPKPSPTTPKEPAPTTTKSAPTTTKEPAP--TTTKSAPTTKEPSPPTTKE 425
Db 449 PVTKAATTTNATPPPOPIPKAATTTTATPVTPQOPIKACTDAPPAVPAKAPSDORAAT 508
OY 426 P---APTTPKEPAPTTTPKAPPTTKEPAP-----TTKEPAPPT-TTKPAPATAP 471
Db 509 PCVPAATNDPOKPPPTPOVPSAVTEPKPOGRAAPPSNEATPAVPSPNLKSPLPTIP 568
OY 472 K-----EPAPTPK---ETAPTPPKLTPTPEKLAPTTPEKPADTPEELA----- 515
Db 569 KPVPLMALTPQVTAQNTQOLAATKPSPIYKASKP-ALMTPPPPPGLPRAALAAAKILG 627
OY 516 -PTTP-----EPTTPTP---EPAPTPP-----KAAAPNPKP----- 545
Db 628 LPSSPVASAMHAKVMPRIPLAPSPVPMASPASLGDARAVALATNNAASPGAKPPAAGNG 687
OY 546 -----PAPTPPKPEAPTTTKEPAPTTTPKETAPT-----TPKGTAPT----- 581
Db 688 TLMAIPGAANTQMAPLGAAGAQTAPMGAATHVSPMGAGATQMSPGANTHMSPIGA 747
OY 582 -----TIKEPAPTTPKKPAKELAPTTKEP-----TSTSDKAPAPPTP 620
Db 748 GGAATQMSPGAAANTQMSPGATTTQMSPGAAATQPSMGAAATQVATISAGNTMOVSP 807
OY 621 KG--TAPTPKPEAPTTTKEPAPTTTPKGTAPTTKEP--APTTPKPKAPKELAPTTTGP 676
Db 808 MGAATPPQTPSGAATTP-QPSPM-----CAATTLMSPGMAATTPG---PSPGAVTTQPP 859
OY 677 -----TSTSDKP-APTTPKET---APTTPKEP---APTTPKFP-APTTPETPP----- 717
Db 860 PMAATNTTQPPMAASTPQSTPBGAAATTTQSPMGATTTQSPMGASTPQAPPTVAGSPT 919
OY 718 PTTSVSTPTTTPKEPTTHKSPDESTPEL--SAEPTPKALENSPREPGVPTTKTAPATKP 775
Db 920 PPPPIPPSPTAQTSPOQMSSPDPDPKAPSAQAQSPAAHVANASPCGV-TAVSPA--P 975
OY 776 EMITTAQDKTTERDLRTTPETTAAPRKMT-KEATATTEKTEKIKATTTQVSTTTQDP 834
Db 976 IGVTEASPSADGALSPGAATDGPRASPAADVTEAATD--VTAANTAVPA-----EA 1029
OY 835 TPKITTLKTTTLAPKVTYTKITTTTTEIMNKPEETAKPKDRATNSKATTKPKPKPTKAP 894
Db 1030 AP-----TKAKRSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 1077
OY 895 KKPSTKKPKTMPRVKRPKTTTPPKMTSTPELNPISRIAEAMLQTTTRPNQTPNLSKLV 954
Db 1078 -----PAVGDGQOQMTPGAASVPP-----VTEAAVO----- 1104
OY 955 EVNRKSEDAEAGE 969
Db 1105 EAAAAAAGAAERE 1119

```

```

RESULT 10
T25697
Hypothetical protein Fl6f9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid Fl6f9.
A:Reference number: Z20071
A:Accession: T25697
A:Status: preliminary; translated from GB/EMBL/DDAJ
A:Molecule type: DNA
A:Residues: 1-1229 <FOL>
A:Cross-references: EMBL:U67956; PIDN:AA807691.1; GSPDB:GN00028; CESP:Fl6f9.2
A:Experimental source: strain Bristol N2; clone Fl6f9
C:Genetics:

```

A/map position: X
A/Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 671/1; 1001/2; 1122/1

Query Match	9.88;	Score 670;	DB 2;	Length 1229;
Best Local Similarity	27.08;	Pred. No. 1.8e-22;		
Matches 321;	Conservative 98;	Mismatches 382;	Indels 300;	

[illegible]

```

QY 782 KDKTERDRLRTPETTYA-----APK 802
      : || : ||||:
Db 855 ETTTTSAE-----TSETTTSAAFIITGESPENTALOSSOKSEENESSAAEKPCARDFVK 911
      : || : ||||:
QY 803 MKRETAITTEKTEBKITATTO-VIST---TQDTTPFKITTLKTTIAPKVT---TTK 855
      : || : ||||:
Db 912 KKKITVKAETTSAAVASTTTEBITTEKSTLEPPIEATITLNEVTOGPAFVGTGAPUDE 971
      : || : ||||:
QY 856 KITTTELEMNKPEFLAPKDKDATNSKATTTKPO-----KTKAPKPKPT 898
      : || : ||||:
Db 972 TTIINTELELSK-----INNTOISQPKPIDISKTALSLISGLGSTKAPMAPPT 1021
      : || : ||||:
QY 899 STKRPKTPBVRVKPKPTTPTRKMTSTMPLEK-----PISRTAEA 937
      : || : ||||:
Db 1022 I-----HTTDAAFVTAETASINDSGDKIIDEAQPDDEIRRA 1059

```

```

RESULT 11
T45025
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C.Species: Homo sapiens (man)
C.Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C.Accession: T45025
R.Dessèyn, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Lalne, A.
J. Biol. Chem. 272, 3168-3178, 1997
A.Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter
A.Reference number: 222899; MUID:97166151
A.Accession: T45025
A.Status: preliminary; translated from GB/EMBL/DBDJ
A.Molecule type: DNA
A.Residues: 1-3570 <DSS>
A.Cross-references: EMBL:Z72496; NID:g1834502; PIDD:CAA96577.1; PID:g1834503
A.Experimental source: Placenta
C.Genetics:
A.Gene: MUC5B

```

```

Query Match      .. 9.6%  Score 651.5:  DB 2:  Length 3570;
Best Local Similarity 22.7%:  Pred. No. 3..3e-21;
Matches 360;  Conservative 107;  Mismatches 538;  Indels 579;  Gaps 65;

Oy  19  IQQVSEDELCKGKCFESFERGECDDAQ-----CKKYD-----KCCPYESFCAEYADN 69
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db  516 IDQV-GGVILTC-----SLETGLTCKKNEPDGTGRFNNCFYNNRVLCDDY-SHCSSTAT 567
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Oy  70  KKNRTK-----KPPKRPVVDKAGSGLDNDFVY--TPDSTGTOHNKVSISPK 117
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db  568 SSTATPSTPGTWTILKRPPTTATTATTAAGSGATATASSTQATGTHVSTT-----ATPPY 622
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Oy  118 IITAK-----PIPRPSLPPNSDTSKETSILYNNKEITVEE-----KETPTTKQST 164
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db  623 VTSKKAIPSSSPECTATVALPALRSTATTPTATSFATIPSSLSGTTTRLSQTTTPPATVMT 682
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Oy  165 DGEKETTSAKEMOSIEKTA-----KDLAPISKULAKTPP----- 199
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db  683 AEPSSPEPVHSTVLTATTATTGATGSAVADPSSPTGAHTTKVLTLLTPTGFTATPSSSP 742
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Oy  200 -KAET-----TTGPAALT-----TPKEPTPTTP-----KEPATPKPE 231
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db  743 GARLLPWISTTTTPTTRGSGTVTPSSIPDCTHNPVLTLLTIVANGSMATPSSSQTS 802
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Oy  232 PPTPTIKASPTPKBPAPATTTKSAP--TTPKEAPATT--KEPAT-----TPKEPATPK 284
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db  803 GTPSLITTTATTITATGTTSTNMPSSIPGTPPLPVLTITATTPATSSTVVPSALSGLTTH 862
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Oy  285 EPAPTTTKS-----APTP-----KEPATPKPKAPATT-----P 314
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db  863 PPVMTTITTTGRCSLSPSSPHVCTAMTASATSGILGTHHTSESTGSHNPATVTGTOH 922
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Oy  315 KEPAATTPKEPTPTTPKEPATTTKEPATTPKBPAPATPKKPAATTPKEKAPAT--PKPE 372
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db  923 STPALSSHPSSRTTESPSPGTTTPGHTT-----ATSTTTATAPSKTRPTSTLPLSP 976
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

```